Pass-through as an Economic Tool*

E. Glen Weyl† and Michal Fabinger‡

September 2012

Abstract

We extend five principles of tax incidence under perfect competition to a general model of imperfect competition. The principles cover 1) the independence of physical and economic incidence, the 2) qualitative and 3) quantitative manner in which taxes are split between consumers and producers, 4) the determinants of tax pass-through and 5) the integration of local incidence to determine the overall division of surplus. We show how these principles can be used to simplify and generalize the analysis of a range of economic questions such as the optimal procurement of new markets and the welfare effects of third-degree price discrimination.

*This paper previously went under the name "A Restatement of the Theory of Monopoly". While many colleagues have provided valuable and in some cases detailed comments, special debts are due to Kevin Murphy (who originally inspired the work), Luciano de Castro and Michael Salinger (for excellent formal conference discussions), David Atkin and Dave Donaldson (for suggesting new results), Jesse Shapiro and seminar participants at the University of California, Los Angeles (for suggesting ways of reframing the paper) and in particular Jeremy Bulow (for detailed guidance, suggestions of new results and unending encouragement throughout the process of writing and rewriting this paper). Weyl is grateful to the Milton Fund, the Harvard Real Estate Academic Initiative and the Marion Ewing Kauffman Foundation for their financial support, which made possible the excellent research assistance provided by Miguel Espinosa, Ahmed Jaber, Rosen Kralev, Andrew Kramer, Stephanie Lo, Yali Miao, Toan Phan and William Weingarten. All errors are our own.

†Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637: weyl@uchicago.edu.

‡Department of Economics, Pennsylvania State University, University Park, PA 16802: fabinger@psu.edu.

> (W)e may prepare the way for using, as we go, illustrations drawn from the incidence of taxation to throw side-lights on the problem of value. For indeed a great part of economic science is occupied with the diffusion throughout the community of economic changes which primarily affect some particular branch of production or consumption; and there is scarcely any economic principle which cannot be aptly illustrated by a discussion of the shifting of the effects of some tax...

–Alfred Marshall, "Principles of Economics", Book V, Chapter IX

1 Introduction

Following Marshall (1890), standard treatments of a range of topics in perfectly competitive markets are typically taught and analyzed in relationship to tax incidence. For example, Chetty (2009) surveys recent work in public finance that builds on the fact that incidence is often a "sufficient statistic" for various welfare analysis to reduce the number of structural assumptions needed to reach welfare conclusions. Virtually all of this work, however, assumes perfect competition, while much of contemporary economic analysis assumes firms have market power. In this paper we show how the principles of incidence and their use as an analytic tool extends to imperfectly competitive models. We survey, where possible, and extend, where necessary, five fundamental principles of tax incidence under perfect competition to successively more general imperfectly competitive settings: monopoly, symmetric imperfect competition and finally general imperfectly competitive models. We then apply these to economic problems ranging from the effects of third-degree price discrimination to welfare analysis in behavioral models. We use these to show how the logic of incidence can both be used to simplify and unify the exposition of existing analyses, as well as to establish new or generalize existing applied results.

It is useful to consider one example of such a result to see the usefulness of incidence reasoning. Suppose, as we do in Subsection 6.1, that an authority can create a market and wants to select the provider(s) of a concession to maximize the social surplus this creates. In addition to purely public settings (such as spectrum allocation or international trade infrastructure), this sort of problem arises when platforms, such as supermarkets (Armstrong and Zhou, 2011) or websites (Gomes, 2012), allow product sellers to display their wares prominently in exchange for payment because the platform can often capture consumer surplus through an up-front fee (Armstrong, 1999; Bakos and Brynjolfsson, 1999). Dasgupta and Maskin (2000) describe an auction that quite generally screens different arrangements for the profits they create, but does not directly screen consumer surplus. When is this sufficient to achieve the socially best possible choice of providers?

Clearly a sufficient condition is that all arrangements have the same ratio of consumer, and thus social, to producer surplus. This is one of the focal incidence quantities we analyze below. Because

we consider this quantity over a wide range of settings our logic simultaneously implies that, for example,

- if different perfectly competitive arrangements are being considered, it is the (appropriately averaged) ratio of the demand to the supply elasticities that would have to be homogeneous across offerings;

- if different constant marginal cost monopolists were being considered, demand curves would have to have the same (average) curvature and in particular it would be sufficient for them to be linked by the Weyl and Tirole (2012) stretch parameterization of demand;

- if different symmetric oligopoly settings were being considered then the degree of "competitiveness" ($\frac{1}{n}$ in the Cournot model or $1 - \sum_{j \neq i} \frac{\partial q_j}{\partial p_i} \Big/ \frac{\partial q_i}{\partial p_i}$ in the differentiated products Nash-in-Prices model) would need to be homogeneous across arrangements, though one arrangement might, for example, involve monopolistic competition and another homogeneous products with conjectural variations.

Only the details of which terms a simple incidence formula may vary changes in order to establish these various results. Furthermore, as we discuss in Section 6, because many applied problems of this kind all depend on similar principles of incidence, they can be treated simultaneously rather than separately. We begin in Section 2 by largely recapitulating these principles of incidence under perfect competition, framing them in terms of five fundamental points:

1. *Economic v. physical incidence*: Welfare effects of taxation are independent of who physically pays the tax.

2. *Split of tax burden*: Revenue raised (under perfect competition) is paid for by reductions in welfare split between the two sides of the market.

3. *Local incidence formula*: The ratio of the tax borne paid by consumers to that borne by producers, the *incidence*, I equals $\frac{\rho}{1-\rho}$, where the *pass-through rate* ρ is the rise in price to consumers for each infinitesimal unit of specific tax imposed.

4. *Pass-through*: ρ, in turn, is equal to $\frac{1}{1+\frac{\epsilon_D}{\epsilon_S}}$, where ϵ_D and ϵ_S are respectively the elasticities of demand and supply.

5. *Global incidence*: The incidence of a finite (i.e. non-infintesimal) tax change is obtained by replacing the pass-through rate by its quantity-weighted average over the range of the tax change. In particular, the ratio of total consumer and producer surplus discussed above (the incidence of a tax that eliminates the market entirely) equals the quantity-weighted average of the pass-through rate over taxes ranging from 0 to the minimum tax that kills the market.

We then consider successively more general models of imperfect competition and study how each of the principles extends or needs to be modified. We begin with monopoly in Section 3 and note four principal modifications. First, even infinitesimal taxes are now more than fully borne by the two sides as there is a deadweight burden arising from the monopoly's distortion of prices. Second, by the envelope theorem, the monopoly fully bears the tax while the consumer still bears ρ per unit quantity, so $I = \rho$ now. Thus ρ now quantifies the deadweight burden. Building on this, we observe, third, that under constant marginal cost every unit of quantity brought in to compete in the market now has a ratio of deadweight loss reduction to profit reduction, or *social incidence*, $SI = \rho$. This, too, can be integrated up to relate the ratio of deadweight loss to producer surplus to the (now mark-up weighted) average pass-through.[1] Fourth, pass-through now depends not only on the relative elasticity of supply and demand, but also on the curvature of demand.

Next, in Section 4, we consider a general model of symmetric imperfect competition in which firms set the elasticity adjusted Lerner index $\frac{p-mc}{p}\epsilon_D = \theta$, a *conduct parameter*. The conduct parameter is equal to 1 under monopoly, 0 under perfect competition and is typically greater than 1 when firms non-cooperatively price complementary goods. We show how this nests many standard symmetric models of imperfect competition including homogeneous products with symmetric conjectures, symmetrically differentiated products Nash-in-prices competition and monopolistic competition.

Symmetric imperfect competition largely interpolates between the behavior under perfect competition and that under monopoly. In particular $I = \frac{\rho}{1-(1-\theta)\rho}$ and $SI = \frac{\theta\rho}{1+(1-\theta)\rho}$. The formula for pass-through simply interpolates between the monopoly and competition formulas in the case that θ is independent of market conditions, as is the case under both the Cournot competition and complements model, and under the (quasi-linear version of) the Dixit and Stiglitz (1977) model of monopolistic competition. However, two new elements emerge if θ depends on market conditions. First, if, as is typically the case with differentiated products Nash-in-prices competition derived from discrete choice behavior among consumers, θ rises as prices rise/quantities fall, then pass-through is higher than the interpolation suggests. Second, when taking averages for non-infintesimal or global incidence, θ also must be averaged over the relevant range.

Finally in Section 5 we consider our most general model with asymmetric firms, allowing for taxes or exogenous competition that fall heterogeneously on firms. Because of the notational complexity introduced by asymmetries we do not extend all of our principles and instead focus on presenting the model and deriving the local and global forms of (possibly heterogeneous-across-firms) tax incidence.[2] Each firm now has its own idiosyncratic conduct parameter θ_i defined as the ratio of the real (mark-ups times changes in quantity) to the pecuniary (quantities times changes in price)

[1]Throughout the paper, we use the term "mark-up" to refer to the absolute mark-up $p - mc$ rather than the relative mark-up $\frac{p-mc}{p}$ to which it is sometimes used to refer.

[2]Extensions of the other principles are available on request.

induced by the firm's changing its quantity. In the symmetric case, as well as some asymmetric cases such as the Melitz (2003) model, this reduces to the more standard common conduct parameter θ.

However, in the more typical case that θ is heterogeneous across firms, incidence now depends on the covariance between various variables. We use the independence of physical and economic incidence to characterize the effects of taxes in terms of the firms they induce to reduce quantities, saying that the tax "falls on" the firms induced by the tax to reduce quantities. While incidence on consumers depends only on the (quantity-weighted across firms) average pass-through rate, the incidence of taxes on firms is now heavier 1) the more the tax falls on firms with high θ_i, 2) the more it falls on firms with high pass-through and 3) the more pass-through covaries with θ_i across firms. Firms and all private individuals may actually benefit from taxes targeted at firms with negative θ_i. Such negative values of θ_i are typical, for example, among inferior products in the Shaked and Sutton (1982) model of vertically differentiated products competing in prices. These results have natural implications for global incidence: for example, firms gain a greater share of surplus to the extent that it is the large firms which have highest θ_i.

We then discuss, in Section 6, a number of other applications where the logic of incidence guides, simplifies, unifies and/or generalizes the analysis. In a supply chain or regulatory relationship with an imperfectly competitive industry, pass-through is central to optimal policy. The welfare effects of price discrimination are largely determined by comparing incidence properties in the two markets separated by discrimination. Strategic complements v. substitutes, which Bulow, Geanakoplos, and Klemperer (1985b) show determine a wide range of effects in oligopoly theory, are often equivalent to comparisons of pass-through or incidence to simple thresholds. We also briefly allude to similar characterizations of many issues in the analysis of platforms, mergers, product design, behavioral welfare analysis, demand systems and the empirics of international trade. Such characterizations were derived in previous versions of this paper but are omitted for the sake of brevity.

Section 7 concludes by discussing potential directions for future research. An extension of our social incidence results beyond the case of constant marginal cost is included as an appendix.

2 Perfect Competition

We begin by reviewing the analysis of incidence in a perfectly competitive market and showing how this can be extended by integration from local formulas to provide a global characterization. To facilitate comparison between the different settings we consider, we articulate the analysis in terms of five principles, reconsidered successively in each setting. Throughout we assume for simplicity that demand and supply are smooth and that excess supply declines in price so there is a unique equilibrium.[3] We also assume, for the sake of simplicity and consistency with nearly all literature we

[3]However, essentially all results can be extended to the case when these fail. Some of these extensions were included in previous versions of this paper. Details are available on request.

are aware of, that all goods outside the industry of interest are supplied perfectly competitively and thus the welfare of producers arising from consumer substitution to these goods may be ignored. The biases introduced by this assumption, which will become apparent in Section 5, are discussed explicitly in the conclusion, Section 7.

The first and most basic principle of incidence, due to Jenkin (1871–1872) is that of *physical neutrality*: it is irrelevant to the economic incidence of a tax whether it is physically paid by producers or consumers. Let p be the price paid by consumers, $\hat{p}$ be the nominal transaction price, D be the quantity purchased as a function of the price paid by the buyers and S be the quantity supplied as a function of price received by the sellers. In the case that the buyers pay the tax, equilibrium is given by

$$D\left(p\right) = S\left(\hat{p}\right)$$

where $\hat{p} = p - t$, while if the sellers pay the tax equilibrium is given by

$$D\left(p\right) = S\left(\hat{p} - t\right)$$

where $p = \hat{p}$ in this case. In either case, equilibrium is $D\left(p\right) = S\left(p - t\right)$. Thus the equilibrium is identical. We therefore refer from now on to the *pass-through rate* $\rho \equiv \frac{dp}{dt}$ as the rate that prices paid by consumers rises when the tax increases. Implicit to this formalism is an understanding that the amount the nominal price falls when a tax is levied on consumers is $1 - \rho$.

Principle of Incidence (Perfect Competition) 1: *Physical incidence of taxes is neutral in the sense that a tax levied on consumers, or a unit parallel downward shift in consumer inverse demand, causes nominal prices to consumers to fall by $1 - \rho$, where the* pass-through rate $\rho \equiv \frac{dp}{dt}$ *is the rate at which prices to consumers rise when a tax is imposed on producers.*

Under perfect competition, both consumers and producers take prices as given and choose quantities so as to maximize their welfare. Thus, as first argued by Dupuit (1844), if we let $CS(p) = \int_p^\infty D(x)dx$ denote the surplus of consumers, $PS(p-t) = \int_0^{p-t} S(x)dx$ that of producers and Q denote the equilibrium quantity, $\frac{dCS}{dt} = -\rho Q$, while $\frac{dPS}{dt} = -(1-\rho)Q$. Therefore the total tax burden, equal to the equilibrium quantity, is split among consumers and producers with weights ρ and $1 - \rho$. Thus Jenkin observed two further principles:

Principle of Incidence (Perfect Competition) 2: *The total burden of the (infinitesimal) tax is shared between consumers and producers.*

Principle of Incidence (Perfect Competition) 3: *The* economic incidence *(or incidence for short) of the (infinitesimal) tax, the ratio of what is borne by consumers to that borne by producers, is given by $I = \frac{\rho}{1-\rho}$.*

Jenkin next asked what economic factors determined ρ. By the implicit function theorem, $D(p) = S(p-t)$ implies, assuming we begin at 0 tax, that

$$D'(p)\rho = S'(p-t)(\rho - 1) \implies (S' - D')\rho = S' \implies \rho = \frac{S'}{S' - D'} = \frac{1}{1 + \frac{\epsilon_D}{\epsilon_S}},$$

where $\epsilon_D \equiv -\frac{D'p}{Q}$ is the *elasticity of demand* and $\epsilon_S \equiv \frac{S'p}{Q}$ is the *elasticity of supply*. Thus, as the classic result goes, the inelastic side of the market bears the burden of taxation.

Principle of Incidence (Perfect Competition) 4: *The pass-through rate is determined by the relative elasticity of supply and demand,* $\rho = \frac{1}{1+\frac{\epsilon_D}{\epsilon_S}}$. *The pass-through increases in the ratio of the elasticity of supply relative to that of demand.*

Not discussed in Jenkin's analysis, nor in the following literature as far as we have been able to determine, is how to integrate up these local changes into finite or global changes in taxes. Suppose that a tax increases by a finite amount from t_0 to $t_1 > t_0$. Letting $Q(t)$ be the equilibrium quantity in the market as a function of the tax and $\rho(t)$ be the pass-through rate as a function of t, the local analysis implies that

$$\Delta CS_{t_0}^{t_1} = -\int_{t_0}^{t_1} \rho(t)Q(t)dt \text{ and } \Delta PS_{t_0}^{t_1} = -\int_{t_0}^{t_1} [1 - \rho(t)]\, Q(t)dt.$$

Now define the *quantity-weighted average pass-through rate between* t_0 *and* t_1 to be

$$\overline{\rho}_{t_0}^{t_1} \equiv \frac{\int_{t_0}^{t_1} \rho(t)Q(t)dt}{\int_{t_0}^{t_1} Q(t)dt}.$$

Then defining the *incidence between* t_0 *and* t_1 to be $I_{t_0}^{t_1} \equiv \frac{\Delta CS_{t_0}^{t_1}}{\Delta PS_{t_0}^{t_1}}$, we have

$$\overline{I}_{t_0}^{t_1} = \frac{\Delta CS_{t_0}^{t_1}}{\Delta PS_{t_0}^{t_1}} = \frac{-\int_{t_0}^{t_1} \rho(t)Q(t)dt}{-\int_{t_0}^{t_1} [1-\rho(t)]\, Q(t)dt} = \frac{\overline{\rho}_{t_0}^{t_1}}{1 - \overline{\rho}_{t_0}^{t_1}} \frac{\int_{t_0}^{t_1} Q(t)dt}{\int_{t_0}^{t_1} Q(t)dt} = \frac{\overline{\rho}_{t_0}^{t_1}}{1 - \overline{\rho}_{t_0}^{t_1}}.$$

Thus the formula for the incidence of a finite tax change is the same as that for a local tax change where the pass-through rate is replaced by the quantity-weighted average pass-through rate over the range of the finite change. One non-infintesimal change of common interest is that of raising the tax so high as to eliminate the market. Let $\bar{t}$ by the smallest tax, possibly infinite, at which $Q\left(\bar{t}\right)$ is 0. We can call the *average quantity-weighted pass-through rate* $\overline{\rho} \equiv \overline{\rho}_0^{\bar{t}}$. Then the global incidence in the market, $\overline{I} = \frac{CS}{PS} = \frac{\overline{\rho}}{1-\overline{\rho}}$.

Principle of Incidence (Perfect Competition) 5: *The incidence of a finite tax change is the same as that of a local change where the local pass-through rate is replaced by the quantity-weighted*

average pass-through rate over the range of the change. The global incidence of the market, the ratio of consumer to producer surplus, is the quantity-weighted average pass-through rate between no tax and the smallest (perhaps infinite) tax that chokes off all trade.

This implies that the same factors that impact local incidence, relative elasticities of supply and demand, determine the global division of surplus. If demand is more elastic than supply globally, the surplus of the market's existence will accrue primarily to suppliers and conversely. Intuitively, to the extent that pass-through does not vary much as tax rates change, taxing a market hurts most that side of the market that benefits most from its existence. In this case, the fact that one side bears little of the cost of taxation indicates that it gains little from the market existing.

3 Monopoly

Analogously to the perfect competition case, we assume the monopolist's profit function is concave in quantity and that her cost $c(q)$ and demand function, represented below by inverse demand $p(q)$ are smooth. The monopolist's revenues are $p(q)q$ with corresponding marginal revenue $mr(q) = p(q) + p'(q)q$ and her marginal cost $mc(q) = c'(q)$. Note that a (per-unit) tax on consumers does not impact p' or q and simply reduces p by the amount of the tax, while a tax on producers raises marginal cost by t uniformly. The monopolist chooses quantity to equate $mr(q) = mc(q)$. Thus, by the exact same argument as in the competitive case, the first principle of incidence extends to monopoly. To our knowledge this result is due to Jeremy Bulow and, while straightforward, has not appeared in print.

Principle of Incidence (Monopoly) 1: *Physical incidence of taxes is neutral in the sense that a tax levied on consumers, or a unit parallel downward shift in consumer inverse demand, causes nominal prices to consumers to fall by* $1 - \rho$.

As far as we know, the logic of local incidence, Principles 2 and 3, under monopoly is not discussed in previous literature. Because under monopoly consumer behavior obeys the same demand curve as under perfect competition, it continues to be the case that $\frac{dCS}{dt} = -\rho q$. However, a monopolist, unlike a price taker, chooses the market price to maximize producer surplus. Thus, by the envelope theorem, we may take the quantity sold as given in computing the impact of a change in taxes on profits under monopoly. Producer surplus, the monopolist's profits, is $[p(q) - t]\, q - c(q)$ and thus, using the envelope theorem, $\frac{dPS}{dt} = -q$.

We thus obtain two results, one qualitative, the other quantitative. Qualitatively, the burden of a tax is no longer simply shared between consumers and producers. Instead, the total burden on consumers and producers is greater than the revenue raised because the quantity sold is already distorted downward. While this qualitative point is likely widely understood, we do not know an

explicit statement of it as such in the literature and it is far from universally applied.[4]

Principle of Incidence (Monopoly) 2: *The total burden of the tax is more than fully shared by consumers and producers. While the monopolist fully pays the tax out of her welfare, consumers also bear an excess burden.*

Quantitatively the size of the excess burden of the tax, per unit revenue raised, is ρ and is borne by the consumers. Another way of expressing the relationship of pass-through to excess burden, which is useful in quantifying changes in social rather than just consumer welfare, is to consider an alternative shock to the market. For this we specialize to the case when the firm has constant marginal cost c; similar but somewhat more complicated results relaxing this assumption are developed in an appendix following the main text. Consider the exogenous entrance into the market of a quantity of the good, $\tilde{q}$, assumed to be produced at cost c. If we allow q to continue to denote the total quantity sold in the market, the firm's profits are now $[p(q) - c]\,(q - \tilde{q})$ and its marginal revenue is now $p(q) + p'(q)\,(q - \tilde{q})$. Note that increasing $\tilde{q}$ has the same effect on the firm's incentives as changing costs by $p'(q)$:

$$\frac{dq}{d\tilde{q}} = p'(q)\frac{dq}{dt} = p'(q)\frac{\rho}{p'(q)} = \rho.$$

Thus the effect of increasing $\tilde{q}$ on the equilibrium quantity supplied in the market is ρ. Deadweight loss from monopoly is $DWL(q) = \int_q^{q^{\star\star}} m(x)dx$ where q is the current (monopoly) quantity, $q^{\star\star}$ is the socially optimal quantity such that $m\left(q^{\star\star}\right) = 0$ and $m(q) \equiv p(q) - c$. Thus $\frac{dDWL(q)}{dq} = -m(q)$ and the change in deadweight loss that occurs as a result of an increase in $\tilde{q}$ is $\frac{dDWL}{d\tilde{q}} = -\rho m$. On the other hand, the reduction in producer surplus from an increase in $\tilde{q}$ can again by computed by the envelope theorem by holding the optimally chosen q fixed: $\frac{dPS}{d\tilde{q}} = -m$. Thus we have that the *social incidence of competition* $SI \equiv \frac{\frac{dDWL}{d\tilde{q}}}{\frac{dPS}{d\tilde{q}}} = \rho$.

Principle of Incidence (Monopoly) 3: *The incidence of the tax is $I = \rho$. If costs are linear (constant marginal) then the* social incidence of competition $SI \equiv \frac{\frac{dDWL}{d\tilde{q}}}{\frac{dPS}{d\tilde{q}}}$ *is also equal to ρ.*

The analysis of pass-through under monopoly is originally due to Cournot (1838), but our exposition here is more similar to the Marshallian treatment of Bulow and Pfleiderer (1983).[5] Monopoly optimization is given by $mr(q) = mc(q) + t$ and thus

$$mr'\frac{dq}{dt} = mc'\frac{dq}{dt} + 1 \implies \frac{dq}{dt} = \frac{1}{mr' - mc'} \implies \rho = \frac{dp}{dt} = p'\frac{dq}{dt} = \frac{p'}{mr' - mc'}.$$

[4] For example, while Turner (2012) uses the appropriate monopolistic formula for pass-through she uses the perfectly competitive incidence logic in discussing how the gains of subsidies are "split" between the two sides.

[5] However, our discussion and interpretation is largely our own.

Marginal revenue $mr = p + p'q$, which has two terms: price and the negative of the marginal consumer surplus ms that consumers earn when quantity expands. We can thus write

$$\rho = \frac{1}{\frac{p'-ms'}{p'} - \frac{mc'}{p'}} = \frac{1}{1 - \frac{ms'q \cdot p}{q \cdot ms \cdot p'}\frac{q \cdot ms}{q \cdot p} - \frac{mc'q \cdot p}{p'q \cdot mc}\frac{q \cdot mc}{q \cdot p}} = \frac{1}{1 + \frac{\epsilon_D}{\epsilon_{ms}}\frac{ms}{p} + \frac{\epsilon_D}{\epsilon_S}\frac{mc}{p}},$$

where ϵ_S is the elasticity of the inverse marginal cost curve ("the supply function") and $\epsilon_{ms} = \frac{ms}{ms'q}$ is the elasticity of the inverse marginal surplus function. $\frac{ms}{p} = -\frac{p'q}{p} = \frac{1}{\epsilon_D}$ and by Lerner (1934)'s rule

$$\frac{p - mc}{p} = \frac{1}{\epsilon_D} \implies \frac{mc}{p} = \frac{\epsilon_D - 1}{\epsilon_D}.$$

This yields

$$\rho = \frac{1}{1 + \frac{\epsilon_D - 1}{\epsilon_S} + \frac{1}{\epsilon_{ms}}}.$$

Two changes from the competitive formula appear. First, ϵ_D has been replaced by $\epsilon_D - 1$. Under monopoly the elasticity of demand is never below unity and thus, intuitively, the appropriate elasticity of demand is relative to unity, rather than 0. However, this neither qualitatively changes any effects (as $\epsilon_D > 1$ always at the optimum) nor does it introduce any new determinants of pass-through.

Thus the second change, the new inverse elasticity of marginal surplus term, is conceptually more significant. As we discuss more extensively with examples in work in progress (Fabinger and Weyl, 2012), ϵ_{ms} measures the curvature of the logarithm of demand because $(\log D)' = \frac{D'}{D} = \frac{1}{p'q} = -\frac{1}{ms}$ so

$$(\log D)'' = \frac{ms'}{ms^2}\frac{1}{p'} = -\left(\frac{1}{\epsilon_{ms}}\right)\left(\frac{1}{ms}\right)\left(-\frac{1}{p'q}\right) = -\frac{1}{\epsilon_{ms}}\frac{1}{ms^2}.$$

Therefore log-concave demand always has $\frac{1}{\epsilon_{ms}} > 0$ and log-convex $\frac{1}{\epsilon_{ms}} < 0$. It can be shown that if demand is concave then $\frac{1}{\epsilon_{ms}} > 1$, while if it is convex $\frac{1}{\epsilon_{ms}} < 1$.[6]

Another way of viewing $\frac{1}{\epsilon_{ms}}$, proposed by Gabaix et al. (2010), is to notice that if α is the standard tail index for the demand, viewed as a probability distribution of consumer values, then $\epsilon_{ms} = -\alpha$. Therefore, for the generalized Pareto/constant pass-through class of demand functions proposed by Bulow and Pfleiderer, which include linear, exponential and constant elasticity as special cases, $\epsilon_{ms} = 1$ for linear, $\frac{1}{\epsilon_{ms}} \to 0$ for exponential and $\epsilon_{ms} = -\epsilon$ for constant elasticity ϵ. A final way to think of $\frac{1}{\epsilon_{ms}}$ is in relationship to risk-aversion: $1 - \frac{1}{\epsilon_{ms}}$ is the Arrow-Pratt measure of relative risk-aversion of the inverse demand function p if this function were to be viewed as a utility function.

Economists have typically viewed ϵ_{ms}, compared to ϵ_D and ϵ_S, as variously difficult to estimate empirically and form intuitions about; see, for example, Farrell and Shapiro (2010a). This atti-

[6] $\frac{1}{\epsilon_{ms}} = \frac{ms'q}{ms} = \frac{(p''q + p')q}{p'q} = 1 + \frac{p''q}{p'}$. Given that $q > 0 > p'$, if $p'' < (>)0$ then this second term is $> (<)0$ and $\frac{1}{\epsilon_{ms}} > (<)1$.

tude strikes us as overly pessimistic. Suppose, for example, that consumer willingness-to-pay was proportional to income. Then $\frac{1}{\epsilon_{ms}}$ corresponds to the well-known curvature properties of income distributions in the segment of the population representing the marginal consumer of the product. Such properties were used by Saez (2001) to calibrate models of optimal income taxation. In particular, $\alpha \in [1.5, 3] \implies \epsilon_{ms} \in [-3, -1.5]$ appears to fit well in the upper tail for most countries and a much less convex distribution (log-normal) appears to fit lower and middle-range incomes. Superior goods will stretch the distribution of willingness-to-pay and thus have a more negative $\frac{1}{\epsilon_{ms}}$ than this income calculation suggests; inferior goods will compress it and lead to a more concave distribution.

Of course this income example is very specific, though commonly used: consumers' willingnesses-to-pay for most products are not simply proportional to income. However, identifying other specific sources of heterogeneity in willingness-to-pay may help with calibration. In general, idiosyncratic, niche products with highly dispersed values to different populations will have a highly negative $\frac{1}{\epsilon_{ms}}$ while utilitarian products that save a constant amount of money to all will have compressed value distributions and thus higher, likely positive $\frac{1}{\epsilon_{ms}}$.

Principle of Incidence (Monopoly) 4: $\rho = \frac{1}{1+\frac{\epsilon_D - 1}{\epsilon_S}+\frac{1}{\epsilon_{ms}}}$. *The general features of pass-through under competition carry over, but under monopoly it is also the case that the more positively (log-)curved demand is, the higher is pass-through.*

Precisely the same logic as under perfect competition allows the extension of local to finite and global incidence, using the appropriate monopolistic incidence formulas. The logic may also be extended to calculate finite changes in deadweight loss using social incidence. Because the arguments are analogous, we do not repeat them here but simply note that the relevant average pass-through rate is now the *mark-up weighted average pass-through* taken over values of $\tilde{q}$:

$$\widetilde{\rho}_{\tilde{q}_0}^{\tilde{q}_1} \equiv \frac{\int_{\tilde{q}_0}^{\tilde{q}_1} \rho(\tilde{q}) M(\tilde{q})\, d\tilde{q}}{\int_{\tilde{q}_0}^{\tilde{q}_1} M(\tilde{q})\, d\tilde{q}}$$

and the *mark-up weighted average pass-through* relevant for *global social incidence* $\widetilde{SI} \equiv \frac{DWL}{PS}$ is

$$\widetilde{\rho} \equiv \widetilde{\rho}_0^{q^{\star\star}}.$$

The validity of the global formula follows from noting that when $\tilde{q} = q^{\star\star}$ there is neither any deadweight loss nor any producer surplus.

Principle of Incidence (Monopoly) 5: *Incidence of finite and global tax changes follow the incidence formula where pass-through is replaced by quantity-weighted average pass-through over the appropriate range. If the firm has constant marginal cost, finite social incidence of competition*

similarly follows the local incidence formula where pass-through is replaced by mark-up-weighted average pass-through over the range of competition and for the global incidence formula the range of competition is that between no competition and the socially optimal quantity being produced in competition.

Thus the pass-through rate, ρ, differently averaged, determines both the ratio of consumer to producer surplus and the ratio of deadweight loss to producer surplus. This is intuitive because a monopolist bears the full cost of any tax she faces. Therefore any burden borne by consumers is exactly the excess burden. The pass-through therefore measures the local incidence to consumers, rather than producers, and to society compared to producers. By the same arguments as under competition, global incidence is simply an averaging of local incidence.

4 Symmetric Imperfect Competition

We now consider incidence under symmetric, imperfect competition. There are n firms (where n need not be finite and in fact may be a continuum, in which case we denote the measure of firms as μ). At symmetric quantities, each firm produces a quantity q at cost $c(q)$ and receives prices $p(q)$. Q is the industry quantity, which equals nq if there is a finite number of firms and μq otherwise. Let marginal cost $mc(q) \equiv c'(q)$ and the elasticity of the market demand be $\epsilon_D = -\frac{p}{p'q}$. When, as above, we allow for exogenous competition $\tilde{q}$, ϵ_D denotes the elasticity of demand for goods *by the welfare-relevant firms* $-\frac{p}{p'(q-\tilde{q})}$, not the elasticity of the total demand for the good. In all cases we focus on a unique symmetric equilibrium, either because it is the only equilibrium or because we are interested in changes local to this focal equilibrium.

Rather than specify a particular model of firm interactions, which are not crucial for most of the questions of incidence, we instead follow Genesove and Mullin (1998)'s variation on Bresnahan (1989) in postulating that the elasticity-adjusted Lerner index $\frac{p-mc-t}{p}\epsilon_D$ is set equal to a *conduct parameter* θ. Where it is not explicitly relevant, we do not explicitly write $-t$ and instead take it as implicitly part of mc.

In some cases θ may depend on the total level of production q (not just the $q-\tilde{q}$ produced by the firms under welfare consideration) and thus implicitly on the interventions t and $\tilde{q}$. However we are not aware of any standard models of imperfect competition where θ as defined above is directly impacted by the interventions (taxes or exogenous competition) we consider.[7]

This model nests a surprisingly wide range of forms of imperfect competition, far broader than those considered by Bresnahan or Genesove and Mullin. In fact, we are not aware of any commonly used symmetric models it does not include. To illustrate this, we briefly describe how three very different and broad models fall within this framework: quantity choices with symmetric

[7]A potential exception are models of (implicit) collusion in which the fall in profitability from an increase in taxes might direct impact θ.

conjectural variations in the spirit of Bowley (1924) with Cournot's model of competition as a special case, symmetrically differentiated Nash-in-prices competition or collaboration (complementary monopoly) including as a special case Cournot's model of perfect complements and partial equilibrium monopolistic competition with the quasi-linear version of the Dixit and Stiglitz (1977) as a special case.

1. Homogeneous products oligopoly: This is the model in the spirit of Bresnahan and our analysis here is taken from him. There is a market demand function $P(Q)$ where Q is the total industry quantity, where P is defined by $P(Q) = p\left(\frac{Q}{n}\right)$. Each of n firms chooses q_i assuming that when it changes its quantity each other firm will change their quantity by $\frac{R}{n-1}$ in response; when $R = 0$ we have Cournot competition and when $R = -1$ we have homogeneous product Bertrand competition. Firm i's profits are $P(Q)q - c(q)$ and her first-order condition at symmetric $q = \frac{Q}{n}$ is

$$(1+R)\,P'(Q)q + P - mc(q) \implies \frac{P - mc}{P}\epsilon_D = \frac{1+R}{n} = \theta.$$

   If R is a constant, as in the Cournot, Bertrand and Delipalla and Keen (1992) conjectures models, then θ is also. This is true, also, if the differentiated products demand system is linear as in this case D is constant and equal to the ratio of two linear coefficients.

2. Symmetrically differentiated Nash-in-prices: The quantity $q_i\left(p_i, \mathbf{p}_{-i}\right)$ each symmetric firm sells depends on its own price p_i and all of the other $n-1$ firms. At symmetric prices p, $q(p) \equiv q_i\left(p, \ldots, p, \ldots, p\right)$ for any i is the inverse of $p(q)$. By Lerner's rule on the residual demand of each firm, at a symmetric equilibrium each chooses $\frac{p-mc}{p} = \frac{q}{-p\frac{\partial q_i}{\partial p_i}}$ while $\epsilon_D = -\frac{p\sum_j \frac{\partial q_i}{\partial p_j}}{q}$. Thus the equilibrium is

$$\frac{p - mc}{p}\epsilon_D = \frac{\sum_j \frac{\partial q_i}{\partial p_j}}{\frac{\partial q_i}{\partial p_i}} = 1 + \frac{\sum_{j\neq i} \frac{\partial q_i}{\partial p_j}}{\frac{\partial q_i}{\partial p_i}} = 1 + \frac{\sum_{j\neq i} \frac{\partial q_j}{\partial p_i}}{\frac{\partial q_i}{\partial p_i}} \equiv 1 - D = \theta,$$

   where the third equality follows by symmetry and $D \equiv \frac{\sum_{j\neq i} \frac{\partial q_j}{\partial p_i}}{-\frac{\partial q_i}{\partial p_i}}$ is the *aggregate diversion ratio* (Shapiro, 1996) from any individual firm to the rest of the industry, the fraction of sales lost by one firm when it increases its price that are captured by another firm in the industry.[8] This ratio is always positive, and thus $\theta < 1$, if goods are substitutes, but is actually negative if goods are complements, and thus in this case $\theta > 1$. In Cournot's collaboration model where goods are perfect complements, $\frac{\frac{\partial q_j}{\partial p_i}}{\frac{\partial q_i}{\partial p_i}} = 1$ for each i and j as the quantity of each good

[8]We apologize for the use of D to denote both demand and aggregate diversion. We could determine no plausible alternative consistent with the literature.

sold depends only on the sum of the prices. Therefore $D = -(n-1)$ and $\theta = n$. In this special case, again, θ is independent of q.

3. Monopolistic competition: There is a continuum of measure 1 of symmetric firms selling different varieties of a product. We seek to model monopolistic competition in a single industry, rather than a whole economy, as in Atkeson and Burstein (2008). Thus we assume consumers have quasi-linear utility: $U\left(\int_i u(q_i)\,di\right) - \int_i p_i q_i di$, where u is strictly concave and both u and U are smooth.

   If prices are symmetric, utility maximization given concavity of u requires that quantities q are symmetric across goods prices $p = U'(u(q))\,u'(q)$. We assume that $U'(u(q))\,u'(q)$ strictly declines in q (aggregate inverse demand for the goods is downwards sloping) and thus associate $U'(u(q))\,u'(q)$ with $p(q)$.

   The inverse demand function facing each firm i, when the quantity chosen by all other firms is q, is $U'\left(\int_i u(q)\,di\right) u'(q^i)$ which we denote simply by $U'u'(q_i)$. The optimal quantity is thus given by $p + U'u''q = mc(q)$ or $\frac{p-mc}{p} = \frac{-U'u''q}{p}$. On the other hand, $\epsilon_D\left(-\frac{p}{p'q}\right) = -\frac{p}{q\left[U''(u')^2+U'u''\right]}$. Thus

$$\theta = \frac{p-mc}{p}\epsilon_D = \frac{U'u''}{U''(u')^2+U'u''}.$$

   If U is concave the goods are substitutes and $\theta < 1$; if U is convex the goods are complements and $\theta > 1$. In the quasi-linear version of the Dixit-Stiglitz model, $u(q) = q^\beta$ where $\beta \in (0,1)$ and U is flexible as it must be re-normalized for quasi-linearity. A natural specification, proposed by Atkeson and Burstein with a finite number of firms per industry, is that $U(x) = x^\eta$ with $\eta > 0$, which we now assume.[9] In particular, calculations show that $\theta = \frac{1-\beta}{1-\beta\eta}$; if $\eta \geq \frac{1}{\beta}$ then industry inverse demand is weakly upward sloping, so we rule out this case even though $\eta = \frac{1}{\beta}$ is perhaps the most immediate quasi-linear version of the Dixit-Stiglitz model.[10] Clearly here θ is a constant. When $\eta > (<)1$ so that goods are complements (substitutes) $\theta > (<)1$.

We now proceed with our analysis using this general model. Assuming no direct dependence of θ on the physical incidence of taxes, the same argument as before (applied to the new equilibrium conditions) extend the independence of economic and physical incidence.

[9] It is a little-remarked fact that in the (quasi-linear) Dixit-Stiglitz model firms goods are complements so long as $\eta > 1$ and thus it is Pareto improving for the industry to be monopolized, rather than to be monopolistically competitive. In the Atkeson and Burstein model, η in our notation is $\frac{1-\eta}{1-\rho}$ in their notation; because they have a continuum of industries, the quasi-linearity assumption holds exactly despite the fact that their model in principle involves income effects.

[10] $\frac{U'u''}{U''(u')^2+U'u''} = \frac{\frac{u''}{(u')^2}}{\frac{U''}{U'}+\frac{u''}{(u')^2}} = \frac{\frac{\beta(\beta-1)q^{\beta-2}}{\beta^2 q^{2\beta-2}}}{\frac{\beta(\beta-1)q^{\beta-2}}{\beta^2 q^{2\beta-2}}+\frac{\eta(\eta-1)u^{\eta-2}}{\eta u^{\eta-1}}} = \frac{\frac{\beta-1}{\beta q^\beta}}{\frac{\beta-1}{\beta q^\beta}+\frac{\eta-1}{u}} = \frac{1}{1-\frac{\beta(\eta-1)}{1-\beta}} = \frac{1-\beta}{1-\beta\eta}.$

Principle of Incidence (Symmetric Imperfect Competition) 1: *In the same sense as in our discussion of monopoly and perfect competition, economic incidence under symmetric imperfect competition is independent of physical incidence.*

Again, we can apply the envelope theorem to consumers to state that $\frac{dCS}{dt} = -\rho Q$. No simple envelope theorem applies to firms, however, as they are neither price-takers nor joint profit-maximizing price setters. We must therefore compute the incidence on firms. Symmetric-across-firms profits per-firm are $[p(q) - t]\, q - c(q)$. We have $\rho = \frac{dp}{dt} = p' \frac{dq}{dt}$ so $\frac{dq}{dt} = \frac{\rho}{p'}$. Thus the impact on per-firm producer surplus ps is

$$\frac{dps}{dt} = (\rho - 1)\, q + \frac{\rho}{p'}\,(p - t) - mc\frac{\rho}{p'} = q\left(\rho - 1 - \rho \frac{p - t - mc}{p}\frac{-p}{p'q}\right) = -q\left[1 - \rho\,(1 - \theta)\right].$$

Aggregating across firms, we conclude $\frac{dPS}{dt} = -Q\left[1 - (1 - \theta)\rho\right]$. Note that this formula is a linear combination of the formulas under monopoly and perfect competition, with a weight θ on monopoly and $1-\theta$ on the perfectly competitive case. This extension of the monopoly formula was first derived by Atkin and Donaldson (2012), extending a previous draft of this paper and inspired the current revision.

Principle of Incidence (Symmetric Imperfect Competition) 2: *If $\theta < (>) 1$, firms less than (more than) fully bear the cost of the tax. As long as $\theta > 0$ the tax has an excess burden, as the burden on consumers more than fully completes the burden on producers.*

We can also extend the social incidence, again under the assumption of constant marginal cost c, to imperfect competition using the notion of exogenous competition. Suppose that a per-firm exogenous quantity $\tilde{q}$ enters the market. Following the logic of the monopoly section, firms now perceive industry marginal revenue $p + p'\,(q - \tilde{q})$ and equate $p + \theta p'\,(q - \tilde{q}) = c$. Thus an increase in $\tilde{q}$ is equivalent to a reduction in cost/tax of $-\theta p'$, so $\frac{dq}{d\tilde{q}} = \theta\rho$. As with the monopoly logic, $\frac{dDWL}{dq} = -\left[p(q) - c\right] \equiv -m(q)$. As with the effect of tax on profits, we cannot apply any simple envelope theorem under imperfect competition. Profits per firm are $[p(q) - c]\,[q - \tilde{q}]$ so

$$\frac{dps}{d\tilde{q}} = p'\theta\rho\,(q - \tilde{q}) + m(q)\,(\theta\rho - 1) = m(q)\left[\theta\rho\frac{p'\,(q - \tilde{q})}{p(q)}\frac{p(q)}{p(q) - c} + \theta\rho - 1\right] = -m(q)\left[1 + (1 - \theta)\rho\right].$$

Thus, while $I = \frac{\rho}{1-(1-\theta)\rho}$, $SI = \frac{\theta\rho}{1+(1-\theta)\rho}$. Two things are worth noting:

1. Holding fixed ρ, $\frac{\partial I}{\partial \theta} < 0$, so long as $I > 0$. That is, the less competitive conduct is, the more of taxation is borne by firms relative to consumers. Especially in its global interpretation as below (firms capture a greater share of surplus the less competitive is conduct), this is intuitively obvious. Perhaps slightly less obvious is that this continues to hold (tax burden

continues to fall more on firms) as θ increases past 1. While this result holds fixed the pass-through rate, which, as we will see below, is partly determined by θ, the same comparative statics can be shown to hold so long as θ is independent of q holding fixed more primitive quantities such as elasticities or the underlying demand and supply curves.

2. Again holding fixed ρ, $\frac{\partial SI}{\partial \theta} > 0$, again so long as $SI > 0$. That is, the less competitive conduct is, the more "exogenous competition" $\tilde{q}$ reduces deadweight loss compared to its impact on profits. Again, the global version is perhaps more intuitive, though likely less obvious than that for I: the more competitive an industry is the greater, the ratio of profits to the deadweight loss of market power (even with constant marginal cost and thus zero competitive profits). When $\theta > \frac{1+\rho}{\rho} > 1$, as may be the case with complementary goods, exogenous competition actually benefits producers (as well as society). Furthermore as $\theta \to 0$, the ratio of deadweight loss to profits goes to 0. Again, these results can be shown taking more "primitive" quantities than the pass-through rate as fixed, but this analysis is omitted here.

Principle of Incidence (Symmetric Imperfect Competition) 3: *The incidence of a tax is* $I = \frac{\rho}{1-(1-\theta)\rho}$, *while under constant marginal cost the social incidence of competition is* $SI = \frac{\theta\rho}{1+(1-\theta)\rho}$. *Holding fixed pass-through, incidence of taxation falls more on firms the greater is* θ *and social incidence of competition falls more heavily on deadweight loss than on firms the greater is* θ.

Under imperfect competition, quantity is chosen according to $p(q) - \theta ms(q) = mc(q)$, where ms is the marginal surplus per firm. Following the logic of the monopoly case:

$$\rho = \frac{1}{1 - \frac{d\theta}{dq}\frac{ms}{p'} - \theta\frac{ms' q \cdot p}{q \cdot ms \cdot p'}\frac{q \cdot ms}{q \cdot p} - \frac{mc' q \cdot p}{p' q \cdot mc}\frac{q \cdot mc}{q \cdot p}} = \frac{1}{1 + \frac{d\theta}{dq}q + \theta\frac{\epsilon_D}{\epsilon_{ms}}\frac{ms}{p} + \frac{\epsilon_D}{\epsilon_S}\frac{mc}{p}}.$$

Now $\frac{p-mc}{p} = \frac{\theta}{\epsilon_D}$ so $\frac{mc}{p} = \frac{\epsilon_D - \theta}{\epsilon_D}$. Defining ϵ_θ as $\frac{\theta}{q\frac{d\theta}{dq}}$, we can now write

$$\rho = \frac{1}{1 + \frac{\theta}{\epsilon_\theta} + \frac{\epsilon_D - \theta}{\epsilon_S} + \frac{\theta}{\epsilon_{ms}}}.$$

This formula nests and generalizes both the homogeneous products conjectural analysis of Delipalla and Keen (1992) and the differentiated products Nash-in-Prices analysis of Anderson, de Palma, and Keider (2001), illustrating why they reach essentially the same conclusion. The exception is one term, $\frac{\theta}{\epsilon_\theta}$, to which we return shortly. Otherwise in both cases the denominator of this expression is a linear combination of that of monopoly and perfect competition, with the weight on monopoly being θ. When θ is invariant to changes in q, as was the case for many models discussed above (both of Cournot's models and the Dixit-Stiglitz model, for example), this additional term is absent because $\frac{1}{\epsilon_\theta} = 0$. In these cases, the qualitative results from the monopoly section continue to hold and the relative importance of the distinctively monopolistic factor (the elasticity of marginal surplus)

compared to the competitive factors (the relative elasticities of supply and demand) is determined by θ.

Perhaps more interesting is the case when θ depends on q. In this case if θ rises with q ($\epsilon_\theta > 0$), pass-through is smaller than indicated by the linear combination as higher prices create more competitive conduce thereby offsetting the impetus for a price increase. If θ falls with q ($\epsilon_\theta < 0$), pass-through is greater than indicated by the linear combination as higher prices create less competitive conduct thereby exacerbating the initial impetus for the price increase. We now consider an example of the second case: discrete choice-based models of differentiated products Nash-in-Prices competition.

As analyzed above, under differentiated products Nash-in-prices competition, $\theta = 1 - D$, where D is the aggregate diversion ratio. A standard micro-foundation of differentiated products demand is discrete choice (Anderson, de Palma, and Thisse, 1992): every one of a continuous distribution of consumers buys only one unit of at most one good and the firms in the market compete to make this one sale to as many consumers as possible. To simplify notation, we now focus on the case of duopoly but our argument, which closely follows that of Jaffe and Weyl (2010), can easily be extended to the case of any number of symmetric firms.

There are two firms, 1 and 2, each selling a single good. A given consumer i has utility u_1^i from consuming good 1 and utility u_2^i from consuming good 2; no consumer can consume more than one good. (u_1, u_2) are drawn from a distribution with a smooth, full-support density $f(u_1, u_2)$ that is symmetric in its arguments. At prices (p_1, p_2), individuals purchasing from firm 1 are all those who both earn a higher utility from this than from purchasing nothing ($u_1 > p_1$) and earn a higher utility than purchasing from 2 ($u_1 - p_1 > u_2 - p_2$). The measure of such customers is:

$$q_1(p_1, p_2) = \int_{p_1}^{\infty} \int_{-\infty}^{u_1 - p_1 + p_2} f(u_1, u_2)\, du_2 du_1.$$

Therefore, at symmetric prices $p_1 = p_2$,

$$D = \frac{\frac{\partial q_1}{\partial p_2}}{-\frac{\partial q_1}{\partial p_1}} = \frac{\int_{p_1}^{\infty} f(u_1, u_1 - p_1 + p_2)\, du_1}{\int_{-\infty}^{p_2} f(p_1, u_2)\, du_2 + \int_{p_1}^{\infty} f(u_1, u_1 - p_1 + p_2)\, du_1} = \frac{\int_p^{\infty} f(u_1, u_1)\, du_1}{\underbrace{\int_{-\infty}^{p} f(p, u_2)\, du_2}_{\text{market exiters}} + \underbrace{\int_p^{\infty} f(u_1, u_1)\, du_1}_{\text{product switchers}}}.$$

Taking the derivative with respect to price yields that $\frac{d\theta}{dp} = -\frac{dD}{dp}$ has the same sign as

$$f(p,p)\left(\int_{-\infty}^{p} f(p, u_2)\, du_2 - \int_p^{\infty} f(u_1, u_1)\, du_1\right) - \left(\int_{-\infty}^{p} f_1(p, u_2)\, du_2\right) \int_p^{\infty} f(u_1, u_1)\, du_1.$$

The first term is always positive and arises from the fact, emphasized by Jaffe and Weyl, that when prices are higher goods inside the industry are less attractive than those outside and thus exert less

competitive pressure. The second term, which may be of either sign, results from the curvature of demand in a firm's own price. Demand being weakly concave in own-price is sufficient for the second term to be positive (as it implies the distribution function is weakly increasing) as well and, as a result, sufficient for $\frac{d\theta}{dp} > 0$; thus, for example, if the distribution of valuations is uniform $\frac{d\theta}{dp} > 0$. However, this is far from necessary and calculations omitted here show that for most standard statistical distributions used for discrete choice (independent Normal, Logistic or Type I Extreme Value distributions), $\frac{d\theta}{dp} > 0$ over the full range of values even though the distributions generate convex demand functions over some ranges. Thus it is only a small exaggeration to say that under discrete choice with Nash-in-Prices competition typically $\frac{d\theta}{dp} > 0$ and thus $\frac{d\theta}{dq} < 0$ because, intuitively, as products in the markets' prices rise they exert less competitive pressure. Therefore the linear combination logic will typically understate pass-through as the degree of competitiveness falls as prices rise.

We are not aware of an equally classic example where $\frac{d\theta}{dq} > 0$ and thus do not include one here. However other examples, in both directions, could be derived from the effects of an increase in price on the standard conditions that facilitate or hamper collusion (Stigler, 1964; Ivaldi et al., 2003).

Principle of Incidence (Symmetric Imperfect Competition) 4: $\rho = \frac{1}{1+\frac{\theta}{\epsilon_\theta}+\frac{\epsilon_D-\theta}{\epsilon_S}+\frac{\theta}{\epsilon_{ms}}}$. *The new term,* $\frac{\theta}{\epsilon_{ms}}$*, leads to lower (higher) pass-through when higher prices/lower quantities lead the industry to be more (less) competitive.*

Finally, to extend incidence globally, we must average not just pass-through but also the product of pass-through and θ. We define the relevant weighted-averages $\overline{\theta\rho}_{t_0}^{t_1}, \widetilde{\theta\rho}_{\tilde{q}_0}^{\tilde{q}_1}, \overline{\theta\rho}$, and $\widetilde{\theta\rho}$ implicitly, analogously to the previous sections. Clearly in the cases when θ is independent of q, it may be taken out of the average. For example,

$$\overline{\theta\rho} \equiv \frac{\int_0^\infty \theta(t)\rho(t)Q(t)dt}{\int_0^\infty Q(t)dt},$$

where $\theta(t)$, analogously to the above, is shorthand for $\theta\left(q(t)\right)$. Using this definition and following the example, $\overline{I} = \frac{\overline{\rho}}{1-\overline{\rho}+\overline{\theta\rho}}$. We do not comment here further on the interpretation of these extensions in the hope that the discussion of the comparative statics of the incidence formulas above suffices.[11]

Principle of Incidence (Symmetric Imperfect Competition) 5: *Incidence of finite and global tax and, under constant marginal cost, competition changes follow the incidence formula where 1) pass-through is replaced, respectively, by quantity-weighted and mark-up weighted average pass-through over the appropriate range and the products of pass-through and 2)* θ *is replaced, respectively*

[11]Note that specializing this result to the case of constant marginal cost and Cournot competition provides is a simple proof of Anderson and Renault (2003)'s surplus bounds under Cournot competition and also of the univariate Prékopa (1971)-Borell (1975) theorem relating the log-curvature of densities to those of survival functions.

by the quantity-weighted and mark-up weighted average product of pass-through and θ over the appropriate range.

5 General Model

We now consider the most general model of the paper, allowing for asymmetric, imperfectly competitive firms. This generalizes all of our previous results, but requires significant new concepts and notation. Therefore rather than extend all of the principles to this more general context, we develop an analogy to our previous model and show how this can be used to derive the basic incidence formula and its global analog. The other principles can be developed in an analogous, but more notation-intensive manner; results are available on request. See Reny, Wilkie, and Williams (2012) for an elegant exposition of Principle 4 in the homogeneous products with conjectures case.

Each firm, i, which may be part of a finite or continuous set of firms, produces quantity q_i and earns profits $p_i(\mathbf{q})\, q_i - c_i(q_i)$, where $\mathbf{q}$ is the (possibly infinitely-dimensional) vector of quantities produced by all firms. Both p_i and c_i are assumed to be smooth and in the case when there is a continuum of firms, p is assumed to be representable as an integral of a smooth function over the firms' quantities. In most of the development below we act as if there are a finite number of firms, making clear the continuous analog only when necessary.[12] As above, we again focus on a single equilibrium either because it is unique or because the market remains at a local equilibrium under the interventions we consider.

As with the symmetric imperfectly competitive model, rather than directly modeling firm conduct we follow Bresnahan (1989) in specifying a firm-specific conduct parameter, though as above we define it somewhat differently to allow it to extend more broadly than his concept. Let $\mathbf{p}, \mathbf{q}$ and $\mathbf{m} \equiv \mathbf{p} - \mathbf{mc}$ respectively denote the vector of firm prices, quantities and mark-ups, where $mc_i(q_i) \equiv c_i'(q_i)$. Additionally, assume firms have a single-dimensional strategic variable σ_i that determines its actions; this may be price, quantity or some supply function in the spirit of Klemperer and Meyer (1989). Each firm takes the other firms strategies as fixed when changing its strategy. As we will see below, however, this does not rule out traditional conjectural variations models. We assume that all $\mathbf{q}$ and $\mathbf{p}$ are smooth functions of $\boldsymbol{\sigma}$, which is assumed to live in the real line, and always obey the demand system $\mathbf{q}(\boldsymbol{\sigma}) = \mathbf{q}(\mathbf{p}(\boldsymbol{\sigma}))$. We use d rather then ∂ notation in denoting derivatives of $\mathbf{q}$ and $\mathbf{p}$ with respect to σ_i to capture the fact that these may include conjectural variations in some interpretations.

[12]With asymmetries, a countably infinite analog is also possible, but we are not aware of any standard model that uses it and thus never explicitly employ it.

We then let each firm's *conduct parameter*

$$\theta_i \equiv \frac{\mathbf{m} \cdot \frac{d\mathbf{q}}{d\sigma_i}}{-\mathbf{q} \cdot \frac{d\mathbf{p}}{d\sigma_i}} = \frac{\mathbf{m} \cdot \frac{d\mathbf{q}}{d\sigma_i}}{-\mathbf{q} \cdot \left(\frac{d\mathbf{q}}{d\sigma_i}\frac{\partial \mathbf{p}}{\partial \mathbf{q}}\right)}. \tag{1}$$

As before, when we introduce taxes $\mathbf{t}$, $\theta_i = \frac{(\mathbf{m}-\mathbf{t})\cdot \frac{\partial \mathbf{q}}{\partial \sigma_i}}{-\mathbf{q}\cdot \frac{d\mathbf{p}}{d\sigma_i}}$. We begin by discussing the definition of θ_i, then provide several examples analyzing the value of θ_i in specific models.

The numerator of θ_i is the set of all *real* or *non-pecuniary* effects of firm i's changing its strategy. This includes both the profits she earns by selling more units at her own mark-up and the real (non-pecuniary) externalities she exerts on other firms by altering their quantities. Taking only σ_i as directly manipulable, it is socially optimal to set this numerator, and thus θ_i, to 0. Furthermore note that in a perfectly competitive market, $\theta_i = 0$ for all firms as all firms' mark-ups are equal to 0. In both of these senses, $\theta_i = 0$ corresponds to perfectly competitive conduct. However, note that it *does not* necessarily imply that firm i has zero mark-up. If other firms have positive mark-ups and firm i's increasing quantity reduces the quantity of other firms then firm i would have to have a positive mark-up to satisfy $\theta_i = 0$. For example, if all of firm i's marginal sales were taken from other firms, firm i would have to have the same mark-up as the diversion-weighted average for other firms.

On the other hand, the denominator of θ_i is the *pecuniary* effects that firm i's changing its price has through changing prices. These again include both the firm's own self-benefiting market power ($-q_i \frac{dp_i}{d\sigma_i}$) and the pecuniary externalities the firm creates for other firms. Such effects are irrelevant and should be ignored from a social perspective as, following the classic logic of pecuniary effects, every benefit brought by a price change is offset by a harm to individuals on the other side of the market. However, if firm i were seeking to maximize industry profits by choosing σ_i, she would set $\theta_i = 1$ as the total impact on industry profits of a change in σ_i is $\mathbf{m} \cdot \frac{d\mathbf{q}}{d\sigma_i} + \mathbf{q} \cdot \frac{d\mathbf{p}}{d\sigma_i}$. In this sense, $\theta_i = 1$ corresponds to monopolistic/perfectly collusive conduct and $\theta_i > 1$ implies overweighting pecuniary effects (relative to a monopolist's behavior) and thus acting even less competitively than a monopolist.

As a final general comment, note that if all firms are symmetric, as in the previous section, and $\theta_i = \theta$, then increasing all q_i symmetrically raises symmetric, per-firm quantity by dq and the price by dp satisfying $mdq = \theta q dp$, where m, p and q are all symmetric, per-firm magnitudes. Thus

$$\theta = -\frac{m}{q}\frac{dq}{dp} = -\frac{m}{p}\frac{p\frac{dq}{dp}}{q} = \frac{p - mc}{p}\epsilon_D,$$

our definition of θ from the previous section. Hence our generalized definition of θ_i collapses to our symmetric definition when firms are symmetric.

We now consider asymmetric versions of the examples of the previous section and fit them into

this model.

1. Homogeneous products oligopoly: At the market level, everything is the same as in the previous section, but now each firm may have a different cost function, set of conjectures, and therefore equilibrium quantity. An individual firm's profits are $P(Q)q_i - c_i(q_i)$. Firms' strategies are denoted in units of quantity, in keeping with Telser (1972), but we allow for conjectural variations; see Jaffe and Weyl (2012) for a more detailed exposition. In particular, taking the derivative with respect to σ_i and assuming firm i anticipates a reaction in amount r_i^j from firm j, firm i's first-order condition is

$$P'(Q)\left(1+\sum_{j\neq i} r_i^j\right) q_i + P(q_i) - mc_i(q_i) = 0 \implies \frac{m_i}{-q_i\left(1+\sum_{j\neq i} r_i^j\right)P'} = 1. \qquad (2)$$

First consider the case of Cournot when $r_i^j = 0$ for all $j \neq i$. Then $\frac{d\mathbf{q}}{d\sigma_i}$ is zero except for an entry of 1 in position i and thus the numerator of the left hand side of the final equality (2) is that of θ_i in equation (1). On the other hand, $\mathbf{q}\cdot\frac{d\mathbf{p}}{d\sigma_i} = QP'$ and thus the denominator of the final expression in (2) is equal to firm i's share $s_i \equiv \frac{q_i}{Q}$ multiplied by $-\mathbf{q}\cdot\frac{d\mathbf{p}}{dq_i}$. Thus equation (2) implies that under Cournot $\theta_i = s_i$. Note, again, that this differs from Bresnahan's formulation, as there $\theta_i = 1$ under Cournot competition.[13]

With conjectural variations, $\frac{d\mathbf{q}}{d\sigma_i}$ has r_i^j in entry $j \neq i$ and 1 in its ith entry. Thus $\mathbf{m}\cdot\frac{d\mathbf{q}}{d\sigma_i} = m_i + \sum_{j\neq i} r_i^j m_j$. The derivative $\frac{d\mathbf{p}}{d\sigma_i}$ has in each entry $\left(1+\sum_{j\neq i} r_i^j\right)P'$, so $-\mathbf{q}\cdot\frac{d\mathbf{p}}{d\sigma_i} = -Q\left(1+\sum_{j\neq i} r_i^j\right)P'$. Thus $\theta_i = s_i\left(1+\sum_{j\neq i} r_i^j\frac{m_j}{m_i}\right)$. As anticipated, θ_i is higher if firms have positive conjectures and lower if they have negative conjectures. Conjectures about firms with relatively large mark-ups matter more than those with small mark-ups. The firm acts monopolistically (like a cartel) if $1+\sum_{j\neq i} r_i^j\frac{m_j}{m_i} = \frac{1}{s_i}$. Note that in the case of symmetric mark-ups and quantities this collapses to our formula for θ from above $\frac{1+R}{n}$.

2. Differentiated Nash-in-prices: Now we have $\sigma_i = p_i$ and following the logic of the previous section, each firm sets $m_i = \frac{q_i}{-\frac{\partial q_i}{\partial p_i}}$. We have $\mathbf{m}\cdot\frac{d\mathbf{q}}{d\sigma_i} = \sum_j m_j\frac{\partial q_j}{\partial p_i}$. On the other hand, $\frac{d\mathbf{p}}{d\sigma_i}$ has entries 0 everywhere but at position i where it has entry 1. Thus

$$\theta_i = \frac{\sum_j m_j\frac{\partial q_j}{\partial p_i}}{-q_i} = \frac{m_i - \sum_{j\neq i} d_i^j m_j}{-q_i\frac{1}{\frac{\partial q_i}{\partial p_i}}} = \frac{m_i - \sum_{j\neq i} d_i^j m_j}{m_i} = 1 - \sum_{j\neq i} d_i^j\frac{m_j}{m_i},$$

[13] Note that while the above logic is left un-changed by including a tax, as long as this is included in the definition of m_i, including exogenous competition makes the relevant $s_i = \frac{q_i - \widetilde{q}_i}{Q - \widetilde{Q}}$. Thus increasing exogenous competition causes θ_i to fall for firms whose effective (after exogenous competition) share it causes to fall and rise for firms whose effective share it causes to rise. θ_i also directly depends on interventions in the models that follow. However, given that the derivative of θ_i does not enter into our analysis below, we omit further discussion of this point.

where $d_i^j \equiv -\frac{\frac{\partial q_j}{\partial p_i}}{\frac{\partial q_i}{\partial p_i}}$ is the *diversion ratio between good i and good j* as the effect of an increase in q_i is equivalent to a change in price of size $\frac{1}{\frac{\partial q_i}{\partial p_i}}$. This quantity is familiar from antitrust analysis: $d_i^j m_j$ is Farrell and Shapiro (1900)'s *Upward Pricing Pressure* from good j to good i were the firms to merge.

As in the symmetric case, diversion plays a key role in determining θ, except that now the proper weighting is given by relative mark-ups. Note that θ_i may be negative (a firm may be pricing below the socially optimal level) if it has a small mark-up relative to the firms from which it diverts sales. For example, a firm whose demand is highly elastic to price (because of strong substitution to outside goods) but diverts from other firms whose goods are not highly elastic to price (because of low outside-substitution) may have a negative θ_i.

For instance, suppose there are only two firms. Then for firm 1,

$$\theta_1 = 1 - d_1^2 \frac{m_2}{m_1} = 1 + \frac{\frac{\partial q_2}{\partial p_1}}{\frac{\partial q_1}{\partial p_1}} \frac{q_2}{q_1} \frac{\frac{\partial q_1}{\partial p_1}}{\frac{\partial q_2}{\partial p_2}} = 1 + \frac{\frac{\partial q_1}{\partial p_2}}{\frac{\partial q_2}{\partial p_2}} \frac{q_2}{q_1},$$

where the third equality follows by Slutsky symmetry. $\frac{\partial q_1}{\partial p_2} \Big/ \frac{\partial q_2}{\partial p_2}$ may be arbitrarily close to -1 if firm 2 has very little substitution to the outside good. Thus if q_2 is large relative to q_1, θ_1 may be arbitrarily negative and thus taxes that fall on (cause reductions in the quantity of) firm 1 may be socially desirable despite firm 1 potentially having positive and even large mark-up.

A simple, if extreme, set of situations where this holds is the duopoly case on which Shaked and Sutton (1982) focus in their model of "vertical" product differentiation where firms are characterized by a quality level u_i and consumers are characterized by an income level y. Consumers with income y gain utility $u_i y$ from consuming a product of quality u_i. Suppose there are two products, one with higher utility than the other ($u_2 > u_1$) and that *some* consumer strictly prefers good 1 so that $q_1 > 0$. Then Shaked and Sutton show that any consumer who is indifferent between good 2 and no purchase will strictly prefer purchasing good 1. Thus there is no substitution between good 2 and the outside good; the only substitution out of good 2 is to good 1 and $\frac{\partial q_1}{\partial p_2} \Big/ \frac{\partial q_2}{\partial p_2} = -1$. Thus θ_1 is negative so long as $q_2 > q_1$ which Shaked and Sutton show occurs whenever y is uniformly distributed and the goods have symmetric constant marginal costs, but it can easily be shown this occurs in many, perhaps most, other cases of interest. Intuitively, if firm 1 is a copy-cat that diverts lots of sales from good 2 but which is highly elastic to outside goods because of its low quality, $\theta_1 < 0$. On the other hand if all goods are complements, $\theta_i > 1$ as $d_i^j < 0$ for all i, j.

3. Monopolistic competition: As in the symmetric case there is a measure-1 continuum of firms,

but now each firms' good may have a different (strictly concave) utility index, $u_i(q_i)$. As before, each firm sets $m_i = -U' u_i'' q_i$. In the continuum model $\mathbf{m} \cdot \frac{d\mathbf{q}}{d\sigma_i} = \int_j m_j \frac{dq_j}{d\sigma_i} dj$ and *mutatis mutandis* for other dot products. As in the traditional quantity interpretation of monopolistic competition the quantity is the choice variable, $\frac{dq_j}{dq_i} = 1$ for $j = i$ and 0 otherwise. Thus $\mathbf{m} \cdot \frac{d\mathbf{q}}{d\sigma_i} = m_i di$.[14] We have $\frac{dU'}{d\sigma_i} = U'' u_i' di$ so for $j \neq i$, $\frac{dp_j}{d\sigma_i} = U'' u_i' u_j' di$ and for $j = i$, $\frac{dp_j}{d\sigma_i} = U' u_i''$. Thus $\int_j q_j \frac{dp_j}{d\sigma_i} dj = \left(U'' u_i' \overline{q_j u_j'} + U' u_i'' q_i\right) di$, where $\overline{q_j u_j'}$ is the average value of $q_j u_j'$ over all other firms. Thus

$$\theta_i = \frac{-U' u_i'' q_i di}{-\left(U'' u_i' \overline{q_j u_j'} + U' u_i'' q_i\right) di} = \frac{U' u_i'' q_i}{U'' u_i' \overline{q_j u_j'} + U' u_i'' q_i}.$$

As in the symmetric case, this is $< (>) 1$ if $U'' > (<) 0$. A natural example is the quasi-linear version of Melitz (2003)'s extension of the Dixit-Stiglitz model to firms with heterogeneous but constant marginal costs. In this model each firm has $u_i(q) = q^\beta$ and $c_i(q_i) = \frac{q_i}{\gamma_i}$, where γ_i is a firm's productivity.[15]

We assume, as in the version of the Dixit-Stiglitz model we solved, that $U(x) = x^\eta$. We observe that in terms of $x \equiv \int_j u_j dj = \int_j (q_j)^\beta dj$, the average $\overline{q_j u_j'}$ may be expressed as $\overline{q_j u_j'} = \int_j q_j u_j' dj = \beta \int_j (q_j)^\beta dj = \beta x$. The equation for the conduct parameter then becomes

$$\theta_i = \frac{1}{\beta \frac{U'' x}{U'} \frac{u_i'}{u_i'' q_i} + 1} = \frac{1}{\beta(\eta - 1)\frac{1}{\beta - 1} + 1} = \frac{1 - \beta}{1 - \beta\eta},$$

which is constant common across firms and exactly the same as in the basic Dixit-Stiglitz model. Thus in the quasi-linear Melitz model, all firms have the same conduct parameter as they would in a quasi-linear Dixit-Stiglitz model with the same demand parameters.

With asymmetric firms our definition of the relevant taxes and pass-through rates we consider also must change. We allow now for the tax $\boldsymbol{\tau}$ to fall heterogeneously on firms, assuming by way of normalization that a unit of the tax has total quantity-weighted size 1: $\frac{\boldsymbol{\tau} \cdot \mathbf{q}}{Q} = 1$, where $Q = \mathbf{1} \cdot \mathbf{q}$. The size of the tax imposed is denoted by $t_{\boldsymbol{\tau}}$ so the total tax is $t_{\boldsymbol{\tau}} \boldsymbol{\tau}$.[16] Pass-through is now a vector, dependent on the $\boldsymbol{\tau}$ considered: $\boldsymbol{\rho}_{\boldsymbol{\tau}} \equiv \frac{d\mathbf{p}}{dt_{\boldsymbol{\tau}}}$.

To form a basis for all $\boldsymbol{\tau}$ we assume that either $\frac{d\mathbf{p}}{d\sigma_i}$ or $\frac{d\mathbf{q}}{d\sigma_i}$ are of full rank, as is true in all (non-degenerate) models considered above and all others we are aware of. We define $\boldsymbol{\tau}_i$ to be such that $-\boldsymbol{\rho}_{\boldsymbol{\tau}_i}$ points in the same direction as $\frac{d\mathbf{p}}{d\sigma_i}$ and $-\frac{d\mathbf{q}}{dt_{\boldsymbol{\tau}_i}}$ points in the same direction as $\frac{d\mathbf{q}}{d\sigma_i}$; we use

[14] Our notation here in terms of differentials is not precise or standard from a formal perspective. However it is more convenient and intuitive for readers not familiar with functional analysis. A more formal analysis is available on request.

[15] Note that we drop fixed costs as they are not relevant to the static analysis we perform.

[16] $\boldsymbol{\tau}$ is allowed to vary with the level of t, as it may have to maintain the normalization.

the negative convention as taxes and quantities typically move in opposite directions.[17] By linear independence, any $\boldsymbol{\tau}$ is a linear combination of $\boldsymbol{\tau}_i$. Thus we collect the coefficients of the linear combination and label them $\boldsymbol{\lambda}^{\boldsymbol{\tau}}$. Intuitively, $\boldsymbol{\lambda}^{\boldsymbol{\tau}}$ tells us who the tax falls on not physically but in terms of the induced changes in firms' economic strategies (and thus quantities and prices). This effectively extends Principle of Incidence 1: it states that it is the incidence of the tax depends not on who pays but on the real changes in strategy these induce.

To determine the costs of taxation borne by consumers, we can, as repeatedly above, employ the definition of consumer surplus:

$$\frac{dCS}{dt_{\boldsymbol{\tau}}} = -\boldsymbol{\rho}_{\boldsymbol{\tau}} \cdot \mathbf{q} = -\rho_{\boldsymbol{\tau}} Q.$$

$\rho_{\boldsymbol{\tau}} \equiv \frac{\boldsymbol{\rho}_{\boldsymbol{\tau}}}{Q}$ is the *quantity-weighted average pass-through rate (across firms)*, from which we omit a bar to avoid confusion with the previous quantity weighting over levels of the tax rather than across firms.

For the impact on producer surplus, it is useful to use the decomposition from above (again, in the derivation we assume a discrete number of firms; the extension to a continuum is natural):

$$\frac{dPS}{dt_{\boldsymbol{\tau}}} = \sum_i \lambda_i^{\boldsymbol{\tau}} \frac{dPS}{dt_{\boldsymbol{\tau}_i}} = \sum_i \lambda_i^{\boldsymbol{\tau}} \left[\frac{d\mathbf{p}}{dt_{\tau_i}} \cdot \mathbf{q} + \frac{d\mathbf{q}}{dt_{\tau_i}} \cdot (\mathbf{m} - \mathbf{t}) \right] - Q = \sum_i \lambda_i^{\boldsymbol{\tau}} \left(\boldsymbol{\rho}_{\boldsymbol{\tau}_i} \cdot \mathbf{q} - \theta_i \boldsymbol{\rho}_{\boldsymbol{\tau}_i} \cdot \mathbf{q} \right) - Q =$$

$$Q \left[\sum_i \lambda_i^{\boldsymbol{\tau}} \left(\rho_{\boldsymbol{\tau}_i} - \theta_i \rho_{\boldsymbol{\tau}_i} \right) - 1 \right] = -Q \left[1 - (1-\theta) \rho_{\boldsymbol{\tau}} + \mathrm{Cov}\left(\lambda_i^{\boldsymbol{\tau}} \rho_{\boldsymbol{\tau}_i}, \theta_i \right) \right],$$

where $\theta \equiv \frac{\sum_i \theta_i}{n}$ and $\mathrm{Cov}\left(\lambda_i^{\boldsymbol{\tau}} \rho_{\boldsymbol{\tau}_i}, \theta_i \right)$ represents the covariance between of the product of (quantity-weighted average) pass-through and the targeting of the tax $\lambda_i^{\boldsymbol{\tau}}$ in terms of the firms it causes to reduce quantity, on the one hand, and the conduct parameters θ_i on the other hand.

Note, first, how this expression collapses in the relevant special cases to those we have thus far considered:

1. Perfect competition: If $\theta_i = 0$ for all firms, then $\frac{dCS}{dt_{\boldsymbol{\tau}}} = -\rho_{\tau} Q$ and $\frac{dPS}{dt_{\boldsymbol{\tau}}} = -\left(1 - \rho_{\tau}\right) Q$. Thus we can allow for multiple products and heterogeneously applied taxes by simply replacing the pass-through rate with the quantity-weighted average pass-through rate, as is typically done for perfectly competitive aggregation. This is the reason why under perfect competition we did not use the space to include multiple products or heterogeneously applied taxes: aggregation is too familiar to yield additional insights.

2. Monopoly: If all firms have $\theta_i = 1$ then $\frac{dPS}{dt_{\boldsymbol{\tau}}} = -Q$, just as under monopoly. Thus a perfect

[17] Note that these conditions are always compatible as $\frac{d\mathbf{p}}{dq_i} = \frac{d\mathbf{q}}{dq_i} \frac{\partial \mathbf{p}}{\partial \mathbf{q}}$ and are often equivalent (whenever $\frac{\partial \mathbf{p}}{\partial \mathbf{q}}$ is of full rank). However, under the Cournot models they are not equivalent as $\mathbf{p}$ is uni-dimensional (constant across firms) in the case of competition and $\mathbf{q}$ is uni-dimensional in the case of collaboration.

cartel with multiple products has the same incidence expression (using the quantity weighted average pass-through) as does a monopoly producing a single product, again justifying our neglect of this case. Thus in an analogous way our results here can be extended to multiple, each multi-product firms, though we do not develop this extension here for the sake of brevity.

3. Symmetric oligopoly: If all firms have the same $\theta_i = \theta$, as in the quasi-linear Melitz model, even if firms are otherwise heterogeneous the covariance term drops out and we recover the symmetric oligopoly expression $\frac{dPS}{dt_{\boldsymbol{\tau}}} = -Q\left[1-(1-\theta)\rho_{\boldsymbol{\tau}}\right]$, again weighting by quantities in the pass-through.

The truly novel term is thus the covariance. This states that firms benefit to the extent that taxes tend to fall upon (quantities tend to fall for) firms with small θ_i. Note this controls for any harm to consumers, which depends only on the overall pass-through and thus it is also in society's interest for taxes to fall on firms with low θ_i. These firms have socially relatively undistorted strategies and thus it is less harmful if their prices rise. In fact, because θ_i may be negative, it may actually cause a social gain to impose taxes that have negative θ_i *even when all firms have positive mark-ups*, as in the Shaked and Sutton example above. Targeting firms with low average pass-through $\rho_{\boldsymbol{\tau}_i}$ is also desirable, but perhaps this was more obvious from above and leads to less extreme conclusions. Similarly, the more pass-through is concentrated among firms with low conduct parameters, the smaller is the burden of taxation on firms.

Global interpretations of the covariance logic above are natural. The same principle of quantity-weighted averaging for integration continues to apply and must now also be applied to the covariance term. For global incidence, $\overline{t_{\boldsymbol{\tau}}}$ now must be such that the quantities in all markets, not just the one market, are 0. Industries where firms with the largest quantities have the highest conduct parameters (large θ_i) will require higher taxes on those firms to eliminate all quantity from the market. This will cause a large covariance and thus a heavy incidence on firms. Thus the ratio of consumer to producer surplus will, all else being equal, be small in industries in which large firms have the least competitive conduct. Similarly industries where firms with high pass-through have high conduct parameters will tend to have heavy incidence and thus large producer relative to consumer surplus. These are both intuitive because producer surplus is harmed by a strong representation, either in quantity or weight due to pass-through, of firms with low conduct parameters.

6 Applications

In this section we discuss a few of the many applications of incidence approach, to provide a bit more substance to the claim that the logic of the preceding sections summarizes the relevant factors in a range of economic analyses.

6.1 Procuring new markets

A public authority seeks to select the provider(s) of a concession to maximize the social surplus this creates. Suppose that each of the concession operators will charge a uniform price if she is selected to be among the operators, each (oligopolistic) group of operators have a single potential proposal and the members have private information on both the consumer surplus CS and profits PS this proposal will generate. The authority cannot or is unwilling to monitor prices ex-post to avoid monopoly distortions and thus must simply choose the operator generating most surplus. We discuss relaxing these assumptions at the end of this subsection.

In addition to purely public settings, similar trade-offs arise when platforms, such as supermarkets (Armstrong and Zhou, 2011) or websites (Edelman, Ostrovsky, and Schwarz, 2007), allow product sellers to display their wares or advertisements for these prominently in exchange for payment, because, as Gomes (2012) argues, the platform has an incentive to internalize the consumer surplus generated by these products in order to profit from consumers on other offerings such as fixed fees for using the platform. In these literatures our assumptions of no discrimination, ex-post monitoring or project selection are maintained.

Solving for the optimal mechanism in this multidimensional context is beyond the scope of our analysis here, so instead here we focus on deriving some analytical principles directly from the logic of incidence.[18] To begin, note that social surplus is the sum of consumer surplus and producer surplus, or $\left(1+\overline{I}\right)PS$. Because only PS affects the incentives of the various potential operators to seek the concession, the reasoning of Jehiel and Moldovanu (2001) suggests that it will typically be impossible to use a mechanism to screen for anything other than PS created by various proposals. If the planner views $\overline{I}$ as symmetrically distributed across groups of firms (conditional on PS), then she wishes to select the firm with highest PS if and only if $E\left[\left(1+\overline{I}\right)PS\middle|\,PS\right]$ is ranked in the same way PS is.[19] A grossly sufficient condition for this is that $\overline{I}$ is distributed independently of PS. Clearly if $\overline{I}$ is constant across all competing groups, this is satisfied, implying Armstrong, Vickers, and Zhou (2009)'s result that if concessions are monopolistic and all firms have linear demand and constant returns (which yields constant pass-through of $\frac{1}{2}$) then the ranking of profits and social surplus are identical and a simple auction is optimal. This also directly implies all the bulleted results in the introduction. However, the greater the noise in $\overline{I}$, the less value there will be to ensuring the highest PS is selected relative to other potential goals (such as revenue maximization).

More generally, an auction should perform reasonably well so long as there is not a strong negative correlation between $\overline{I}$ and PS. If such correlation were too strong, the planner might want

[18]For more on two distinctive approaches to multidimensional mechanism design, see Rochet and Stole (2003) and Veiga and Weyl (2012). The logic of the results given here is closely related to that of the latter paper for obvious reasons.

[19]Technically only the ranking of the top proposal matters, but in many circumstances a full ranking is useful for the same reasons as in social choice theory (different alternatives may be available at different times).

to be unresponsive to PS because of the implied adverse selection, randomizing among symmetric proposals. Especially in such cases, the authority would seek information that would allow it to handicap the auction to favor operators with high expected $\overline{I}$. Factors indicated by the above analysis are that proposals with constant or increasing returns should be favored over those with capacity constraints, those generating highly convex demand should be favored over ones with more concave demand, those with low conduct parameters should be favored over those with high conduct parameters and those in which the largest and highest pass-through firms have low conduct parameters should be favored over the reverse.[20] We suspect that principles of incidence would also play an important role in the design of the optimal mechanism.

This importance of incidence is likely to carry over if our assumptions of a single, fixed proposal for each cluster and no ex-post monitoring are relaxed. For example, Weyl and Tirole (2012) show that when rewards for creating a new market can be based only on equilibrium ex-post prices and quantities, the factor that cannot be screened is $\overline{I}$, so again its statistical relationship to factors that can be screened is crucial to optimal policy. Other formulations would allow investments by individuals in changing their projects while maintaining the fact that only PS can be observed or screened ex-post, as in Holmström and Milgrom (1991), or would allow $\overline{I}$ and PS to both be observable ex-post but the set of projects available for proposal would be unobservable, as in Armstrong and Vickers (2010) and Nocke and Whinston (Forthcoming). In the former case, there the more substitutable are investments in improving the unobservable $\overline{I}$ and the observable PS, the less responsive should be the award (or at least profits) to PS. In the latter case, again both the statistical relationship between $\overline{I}$ and PS and ex-ante information about $\overline{I}$ are relevant. Armstrong and Vickers (2010) and Nocke and Whinston (Forthcoming) analyze these in the closely related case of mergers that we discuss in Subsection 6.5, assuming proposals are selected among those permitted to maximize PS. As Armstrong and Vickers (2010) emphasize, minimum social surplus for a project to be acceptable should rise when there is less correlation between $\overline{I}$ and PS or more variance in $\overline{I}$. As Nocke and Whinston (Forthcoming) emphasize, marginal cost reductions (caused by mergers) for firms with higher I should be favored even conditional on social surplus generated to offset the private bias towards selecting based only on PS.[21]

6.2 Supply chains and optimal taxation

A canonical model of supply chains proposed by Spengler (1950) has an "upstream" firm choosing its price, which is then taken by a (downstream) firm that charges prices to consumers. Natural

[20] One application of this logic is Nocke and Whinston (Forthcoming)'s result that marginal cost reductions (through mergers) for small firms in Cournot oligopoly are more desirable than for large firms; this is because they have lower conduct parameters, from the above analysis. Additionally, while we have assumed uniform pricing, if some firms were able to discriminate more effectively than others, discriminatory firms should be penalized (below their willingness-to-pay to enter the market) as they will appropriate a greater fraction of the social surplus they create.

[21] See the previous footnote.

extensions considered by many authors allow for multiple stages in the supply chain and imperfect competition at each stage.[22]

First consider a supply chain consisting of several layers of imperfectly competitive firms supplying a necessary input to a downstream sector, which may then supply end-consumers or another downstream sector. We focus on the case where each layer is symmetrically imperfectly competitive, but again this may be relaxed. There is a symmetric-at-symmetric prices demand $q(p_0)$ for the product. This, combined with a supply side structure, determines an equilibrium pass-through rate ρ_0 of the retailers as a function of the per-unit cost p_1 they are charged by the upstream firms. The upstream firms thus face effective (symmetric-at-symmetric prices) demand $q(p_0(p_1))$ with elasticity $\epsilon_D \rho_0$, where ϵ_D is the direct demand elasticity downstream. Thus, letting θ_1 be the competitiveness of the level 1 sector, equilibrium in a symmetric upstream market is given by

$$\frac{p_1 - mc}{p_1} \epsilon_D = \frac{\theta_1}{\rho_0}.$$

As a result the comparison of mark-ups between the upstream and downstream firms is given by the comparison of $\frac{\theta_1}{\rho_0}$ to θ_0, where θ_0 is the competitiveness parameter downstream. This reasoning continues up the supply chain, with the aggregate pass-through of all levels beneath determining the incentives faced at each level. This implies that the pass-through from the nth to the $(n-1)$th level will depend on derivative of the pass-through from the $(n-1)$th to the nth level and thus on the $(2+n)$th derivative of demand, in principle allowing the identification from mark-up data of very high-order properties of demand, extending the logic of Villas-Boas and Hellerstein (2006). Conversely, if constant pass-through is assumed many of these effects disappear, strong predictions are implied and the model is highly over-identified.

A slightly modified application of this reasoning involves a two-stage chain where the first stage firm (usually interpreted as a national government taxing foreign trade or regulator of an imperfectly competitive industry) internalizes the welfare of final consumers and/or the taxed downstream industry to greater or lesser extents. While the logic of incidence provides a useful framework for any set of weights, we focus on the case considered by Brander and Spencer (1981) and Laffont and Tirole (1993) when consumer surplus is fully internalized and the welfare of the imperfectly competitive firms is entirely neglected.

The government charges a specific tax t. If D is the equilibrium demand of consumers, the marginal loss to consumers of the product is ρD and to the government on infra-marginal tax is

[22] Analogous settings arise when firms sequentially choose how much of a homogeneous good to produce, as in the classic von Stackelberg (1934) model, extended by Anderson and Engers (1992) to the case when this occurs in many stages. The pass-through of quantities at each stage to the final market plays an analogous role to cost pass-through along a supply chain. Details are available on request.

$-t\rho D'$, while the marginal revenue gain to the state is D. Thus the optimum requires

$$1 = \rho \left(1 + \frac{t\epsilon_D}{p}\right) \implies \frac{t^\star}{p} = \frac{1-\rho}{\rho\epsilon_D}.$$

Note that this formula in no way depends on the existence of imperfect competition; it applies equally well to the setting where the foreign firms are perfectly competitive. It thus unifies the monopoly analysis of Brander and Spencer with the classic analysis of terms-of-trade reasons for taxing imports as in Johnson (1953–1954). It may also be easily extended to asymmetric industries. The only difference is that with an imperfectly competitive foreign sector, it is possible that $\rho > 1$ and thus a negative tax (subsidy) on imports may, in principle, be optimal. Thus the two theories are one, at least this far.

However, the externality of a tax on the foreign sellers is strictly greater with conduct parameters above zero: rather than $D(1-\rho)$, burden on the foreign industry is $D[1-(1-\theta)\rho]$. Thus there will be a stronger incentives for international trade agreements to limit such taxes between countries where firms have higher conduct parameters and in models where firms exercise this power, as shown by Ossa (2011). This is essentially the case when the weight on the regulated or foreign industry's welfare is non-zero because trade negotiations lead it to be internalized.

6.3 Third-degree price discrimination

A recent literature has revisited classical questions in the theory of monopoly price discrimination using an approach closely related to that employed here. Aguirre, Cowan, and Vickers (2010) (ACV) return to one of the oldest questions in industrial organization, posed by Pigou (1920): when does explicit third-degree price discrimination by a monopolist raise output and/or welfare?[23]

Consider two markets, strong (S) and weak (W). Absent discrimination, prices are constrained to be identical. With discrimination, prices in S exceed those in W by Δ. ACV propose a natural continuous path from no discrimination to discrimination: we require that $p_S \leq p_W + \delta$. Assume profits in each market π_S and π_W are concave in price. Then for any $\delta \in [0, \Delta]$, the monopolist will choose $p_S = p_W + \delta$. Her first-order condition is thus $\pi_S'(p_W + \delta) + \pi_W'(p_W) = 0$. For $\delta < \Delta$, $\pi_W' < 0 < \pi_S'$, but these both converge to 0 as δ goes to Δ.

A firm facing exogenous quantity $\tilde{q}$ earns profits $[q(p) - \tilde{q}](p-c)$. Her first-order condition is thus $q'(p)(p-c) + q(p) - \tilde{q}$, while the first-order condition in the high market in the price discrimination problem is $q_S'(p)(p-c) + q_S(p) + \pi_W'(p-\delta)$. In effect, the downward pressure on prices from the constraint against discrimination in the low market enters in the same way as exogenous quantity. Moving towards discrimination is therefore equivalent to moving exogenous quantity from the high

[23]While we focus on the social welfare and output effects, Cowan (2010) also uses incidence elegantly to study the effects on consumer surplus. Similar extensions of his logic as those we propose below for output and welfare are possible for consumer surplus.

market to the low market.

Thus ACV show that discrimination leads to higher output if an average of pass-through in the low market exceeds that in the high market. Similarly the change in social welfare in each market from the change in quantity is $\int M dq$ so a comparison of an average of the mark-up times the pass-through over the relevant range in the two markets determines the welfare effect of discrimination. The connections of pass-through to demand curvature make it clear how this result immediately implies the famous prior results of Pigou (1920), Robinson (1933), Schmalensee (1981) and Varian (1985) on the connections between demand curvature and the effects of discrimination.

The logic of incidence can be used to extend both of these results to symmetric imperfect competition. Suppose that, with or without discrimination, each market is governed by the same conduct $\theta = \frac{p-c}{p}\epsilon_D$ where ϵ_D is either independent or pooled depending on whether discrimination is allowed or not. Absent discrimination, $\epsilon_D = \frac{q_S \epsilon_{D_S} + q_W \epsilon_{D_W}}{q_S + q_W}$. Thus

$$\epsilon_D = -\frac{p\left(q_S' + q_W'\right)}{q_S + q_W} = -\frac{p q_S'}{q_S - \frac{q_W' q_S - q_S' q_W}{q_S' + q_W'}} = -\frac{q_W' p}{q_W + \frac{q_W' q_S - q_S' q_W}{q_S' + q_W'}}.$$

Therefore the argument that the prohibition on discrimination acts as equal-and-offsetting exogenous quantity competition in the two markets in an amount $\frac{q_W' q_S - q_S' q_W}{q_S' + q_W'}$ holds generally. Because quantity pass-through in the two markets is $\theta\rho$, if θ is constant in p, precisely the same results, interpreted in terms of averages of pass-through or of demand curvature (as this is a simple transformation of pass-through), hold under imperfect competition. If θ is not constant, then the result must be replaced with averages of $\theta\rho$.

Similarly the result may also be shown to hold when θ is not the same in the two markets. Then if θ is higher in the strong market it is clearly more likely *ceteris paribus*, that discrimination is harmful as it is more likely that averaged over the relevant range $\theta_S \rho_S m_S > \theta_W \rho_w m_W$. This is a generalization of the result of Holmes (1989) who argues that when discrimination is in favor of individuals for whom competition is more intense (in differentiated products Nash-in-prices competition), discrimination is more likely to be harmful because $\theta = 1 - D$ and D, the aggregate diversion ratio, measures the degree of competition for customers within the market.

6.4 Strategic effects

Bulow, Geanakoplos, and Klemperer (1985b) highlight the importance of strategic effects (whether one firm raising or lowering its price or quantity causes others to follow or do the opposite) in a variety of problems in oligopoly theory, particularly strategies to deter entry or affect post-entry competition as studied by Fudenberg and Tirole (1984) and Bulow, Geanakoplos, and Klemperer (1985a). In this subsection we discuss the close relationship between strategic effects, and their global analog, the strategic impact of entry, on the one hand and incidence and pass-through on

the other.

First consider quantities. Suppose one firm increases the quantity it produces of a good that is homogeneous with other goods. This is precisely the same, from the perspective of the other firms who take quantities as given as in Cournot, as an increase in the exogenous competition discussed in Sections 3 and 4. Thus if the remaining firms are symmetric imperfect competitors (or a residual monopolist) whether they respond by increasing or decreasing their own output is determined by the sign of $\theta\rho - 1$ with constant marginal cost and, by the logic of our appendix, by $\left(\theta + \frac{\epsilon_D - \theta}{\epsilon_S}\right)\rho - 1$ more generally. Because Bulow, Geanakoplos, and Klemperer only consider duopoly settings, this strictly generalizes their results. For example, it gives a simple demonstration of why strategic substitutes are more likely under Cournot when the number of firms is large, as $\theta = \frac{1}{n}$ in this case. The appendix gives further results on the sign of the strategic effect in quantities.

Second consider prices. When one firm raises its prices, this raises the willingness of consumers to pay for other products. In many models, the result is a parallel shift upwards in willingness-to-pay. For example, in symmetric oligopoly models with no outside good such as the "Spokes" model of Chen and Riordan (2007) or the random utility model of Perloff and Salop (1985), any individual firm raising its price increases the price of the only "outside option" for the rest of the symmetrically imperfectly competitive industry. This is equivalent to increasing willingness-to-pay for the remaining industry by an identical amount for all individuals. By Principle of Incidence 1 under symmetric imperfect competition, a unit increase in consumer willingness-to-pay will raise prices if and only if $1 - \rho > 0$ or equivalently $\rho < 1$. This explains why Gabaix et al. (2010) find that in the Perloff and Salop (1985) model entry (a new firm entering the market and thus effectively reducing their price from infinity) raises (lowers) prices under constant marginal cost if the distribution of valuations is log-convex (log-concave) given the relationship between pass-through and log-curvature under constant marginal cost discussed extensively in Section 3. However, it holds more broadly and would also apply if costs were allowed to be non-linear. Chen and Riordan (2008) obtain a similar relationship, even while relaxing the assumption of no outside substitution.

Bulow, Geanakoplos, and Klemperer's examples of the importance of such strategic effects are too numerous to consider here. So we just give one example of how the logic of incidence relates to the normative consequences of strategic complementarity and substitutability. Consider, as in Mankiw and Whinston (1986), a firm that enters production of products already existing in an industry. We specialize compared to Mankiw and Whinston by assuming a symmetric industry prior to the entry and constant marginal cost (though both of these assumptions can be relaxed at some expositional costs) but generalize by allowing any symmetric oligopoly setting (including monopolistic competition and/or differentiated products). The new firm's entry leads to some equilibrium production by the new firm which effectively acts as exogenous quantity entering the market. By the logic of Section 4, deadweight loss falls by $m\theta\rho$ (for each unit sold) while the entering firm gains profits m taken from existing firms. Thus the new firm's profits on a small unit

of increased production are greater (less) than her social contribution if and only if, as Mankiw and Whinston find, $\theta\rho < (>)1$, that is there are strategic substitutes (complements) in quantities. By mark-up-weighted averaging and using the logic of Principle of Incidence 5 under symmetric imperfect competition, these results can be extended to discrete entry in more quantitative manner than in Mankiw and Whinston. The intuition behind the result may also be clarified by the incidence logic: entry is excessive (insufficient) to the extent that negative (positive) real non-pecuniary externalities are created by entry by reducing (increasing) the profitable quantities of existing firms.

6.5 Other applications

Previous drafts of this paper discussed a number of other applications in detail as above. Here we mention them only briefly:

1. *Platforms*: In platform industries, ones where individuals' utility from consumption depends on how many other individuals participate on the platform, an important trade-off is often between reducing prices and increasing network benefits. For example, in the Rochet and Tirole (2003) model, policy analysis has focused on whether individuals on one side of the market (say credit card accepting merchants) might benefit from paying a higher prices that subsidize participation on the other side of the market. Both the amount of subsidy that a price rise finances and the relative size of the benefits to infra-marginal individuals from more network benefits compared to the loses from higher prices are closely related to incidence. A number of results in Rochet and Tirole model, and various extensions such as those analyzed by Goos, Van Cayseele, and Willekens (2012), can be characterized parsimoniously in terms of incidence; see the now-defunct working paper Weyl (2009b), as well as Weyl (2009a) and Goos, Van Cayseele, and Willekens (2012) for more details. The welfare economics of other, quite different platform models, such as Becker (1991), also turn on incidence properties; details are available on request.

2. *Mergers*: A long line of work (Werden, 1996; Shapiro, 1996; Farrell and Shapiro, 2010b) has established a close connection between the impact of mergers in differentiated products industries and the effects of changes in cost. These ideas have been incorporated into policy through the new United States and United Kingdom horizontal merger guidelines. They suggest that agency investigators consider the equivalent of a merger in marginal cost changes to determine its competitive effects. Jaffe and Weyl (2012) show that a (matrix) product of pass-through rates and these equivalent cost changes are a first-order approximation to the effect of a merger on prices, where the approximation's error is proportional to the curvature of the demand system and the square of the size of the equivalent cost changes. In a previous

version of this paper we showed a stronger result: that because of the relationship between local pass-through and global incidence, under constant marginal cost and constant pass-through a merger-to-monopoly between two firms with small diversion eliminates a fraction of consumer surplus equal to the diversion ratio between the products. Similar approximations may be used to derive the effect of mergers on profits or deadweight loss, further exploiting the logic of incidence.

3. *Product and market design*: Spence (1975) and Johnson and Myatt (2006) ask how firms choose a non-price characteristic of their product to affect the demand according to a specified, parametric relationship between demand and the characteristic. An alternative approach to the problem of product (or market) design is to assume a firm or producer surplus-maximizing entity can choose any arrangement of demand and supply subject to some constraint. A simple constraint would be that total potential gains from trade are constant. The principles of incidence give simple ways to compare different arrangements subject to this constraint: firms want arrangements where $\overline{I}$ and $\widetilde{SI}$ are both low. This, for example, means monopolists with constant marginal costs prefer demand to be as concave as possible, that perfect competitors prefer inelastic supply and elastic demand, that symmetric competitors prefer θ to be as close to 1 as possible and to increase at higher prices and that asymmetric competitors put a premium on high conduct parameters at large, high-pass-through firms.

4. *Behavioral welfare analysis*: There has been a recent revival of interest, surveyed by Mullainathan, Schwartzstein, and Congdon (Forthcoming), in Dixit and Norman (1978)'s analysis of welfare when consumers may fail to act or be persuaded not to act in their own best. Incidence plays a crucial role in much of this analysis; we consider two examples. Farrell (2008) considers markets in which secret fees may be charged on goods that consumer do not understand. These act as consumer-financed subsidies of the products, but may actually benefit consumers if $I > 1$. Dixit and Norman consider advertising that uniformly raises the willingness of consumers to pay for a product. This acts as a similar subsidy, but Dixit and Norman are concerned with social welfare which is always increased locally by such a subsidy if $\theta > 0$. Instead they ask when and to what extent advertising is excessive. While they derive qualitative conditions to ensure excessive advertising, their analysis can be made more quantitative and extended beyond the monopoly setting they consider using the logic of incidence as this determines how the gains from the subsidy are split between externalities (from the firm's perspective) to consumers and benefits to the firm. Even in rational models of uniform-shift advertising, such as that of Becker and Murphy (1993), incidence plays a crucial role in welfare analysis for similar reasons.

5. *Demand systems*: Demand curvature plays a key role in determining pass-through and thus incidence when firms have market power. While most standard demand forms used when

modeling imperfect competition allow flexibility in the elasticity of demand, few are flexible in the curvature properties of demand, as we show in related work (Fabinger and Weyl, 2012). The restrictions imposed are not typically based on ex-ante economic intuitions about the relevant environment; in fact, different forms that impose different restricts are often used somewhat interchangeably despite generating very different implications. We thus propose a new class of *Adjustable pass-through* (Apt) demand forms that substantially increase the flexibility of curvature while maintaining tractability and nesting the most common demand forms.

6. *An empirical application*: Atkin and Donaldson (2012) explicitly use the role of incidence as a "sufficient statistic" and the structure of our results above to analyze the degree of competition in markets and the division of surplus from globalization. They consider markets for various internationally-traded commodities in different locations within developing countries in South Asia and Sub-Saharan Africa. In a symmetric imperfect competition model they impose three key assumptions: that demand curvature is constant (demand is in the Bulow and Pfleiderer class) and the same across markets for a given product, returns to scale are constant and conduct is invariant to prices (θ is constant). They then use the variation in empirical pass-through in the face of global price shocks across geographic locations for a given product to back out θ. Integrating and using the fact the under their assumptions local and global incidence are identical, they determine the division of surplus arising from the market existing between the intermediaries and consumers.

7 Conclusion

This paper argues that incidence offers just as powerful a framework for organizing the analysis of comparative statics and welfare under imperfect competition as it does under perfect competition. We have argued that, to paraphrase the conclusion of Bulow, Geanakoplos, and Klemperer (1985b), the crucial question for welfare in imperfectly competitive markets is typically not "Do these markets exhibit price competition or quantity competition or competition using some other strategic variable?"; "Are products differentiated, how many firms are there, do firms act strategically or are they monopolistic competitors?" or even Bulow, Geanakoplos, and Klemperer's "Do competitors think of the products as strategic substitutes or as strategic complements?"; but rather, "What is the pass-through and incidence of a tax in this market?" Unlike the first group of questions, this last is not "new" to oligopoly theory. Rather it is what, at least since the time of Marshall, economists have been asking about competitive markets to analyze a wide range of outcomes and policies. And as discussed in Section 6, once incidence and pass-through have been determined, there is, for many questions, little difference between the relevant analysis in perfectly competitive markets and

imperfectly competitive markets. Thus the analysis of "strategic" industries with market power may not be as distinct as it may at first seem from the analysis of perfectly competitive markets.

One of the most important weaknesses of our analysis was that we followed nearly all partial equilibrium literature in assuming that all goods outside the industry under consideration were perfectly competitively supplied (with no externalities). Section 5 made clear how problematic this assumption is: if goods outside the considered industry that are either complementary or substitutable with goods in the industry have positive mark-ups, then firms in the industry having 0 mark-ups is not typically optimal. Considering detailed behavior outside the industry of interest would defeat much of the simplifying value of a partial equilibrium analysis. However, it would be useful to develop a version of the analysis here where it was assumed that all goods outside the market to which consumers might substitute have so fixed average mark-up from the economy, rather than 0 mark-up.

Some additional applications and extensions of the framework seem natural:

- The design and analysis of auctions, some connections to which are drawn by Mares and Swinkels (2011).

- Shifting bargaining power from one side to the other side of the market in the Riley and Samuelson (1981) model of bilateral trade can be shown to have effects similar to changing the amount of exogenous competition facing each side of the market its optimization of price. Thus if the principles of incidence could be extended to such bilateral imperfect competition settings, they might be used to provide an elegant characterization of optimal bargaining power in bilateral trade.

- Nonlinear pricing and optimal taxation, where curvature of distributions was shown by Saez (2001) to play an important role.

- Almost all international trade models use explicit, often constant pass-through demand forms to obtain results, which are known to vary based on, for example, whether linear or constant elasticity demand are employed. It thus seems likely that incidence plays an important role in the comparative statics of such models.

Finally, we assumed throughout most of the paper that firms' only instruments were uniform prices and all consumers were homogenous in their value to firms. Extending the logic of incidence to cases with heterogeneously valuable consumers and non-price, or discriminatory price, instruments as summarized by Veiga and Weyl (2012) is a promising direction for future research.

References

Aguirre, Iñaki, Simon George Cowan, and John Vickers. 2010. "Monopoly Price Discrimination and Demand Curvature." *American Economic Review* 100 (4):1601–1615.

Anderson, Simon P., André de Palma, and Brent Keider. 2001. "Tax Incidence in Differentiated Product Oligopoly." *Journal of Public Economics* 81 (2):173–192.

Anderson, Simon P., André de Palma, and Jacques-François Thisse. 1992. *Discrete Choice Theory of Product Differentaiation*. Cambridge, MA: MIT Press.

Anderson, Simon P. and Maxim Engers. 1992. "Stackelberg Versus Cournot Oligopoly Equilibrium." *International Journal of Industrial Organization* 10 (1):127–135.

Anderson, Simon P. and Régis Renault. 2003. "Efficiency and Surplus Bounds in Cournot Competition." *Journal of Economic Theory* 113 (2):253–264.

Armstrong, Mark. 1999. "Price Discrimination by a Many-Product Firm." *Review of Economic Studies* 66 (1):151–168.

Armstrong, Mark and John Vickers. 2010. "A Model of Delegated Project Choice." *Econometrica* 78 (1):213–244.

Armstrong, Mark, John Vickers, and Jidong Zhou. 2009. "Prominence and Consumer Search." *RAND Journal of Economics* 40 (2):209–233.

Armstrong, Mark and Jidong Zhou. 2011. "Paying for Prominence." *Economic Journal* 121 (556):F368–F395.

Atkeson, Andrew and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review* 98 (5):1998–2031.

Atkin, David and Dave Donaldson. 2012. "Who's Getting Globalized? Intra-national Trade Costs and World Price Pass-Through in South Asia and Sub-Saharan Africa." Https://dl.dropbox.com/u/2734209/Atkin_Donaldson_WGG_paper.pdf.

Bakos, Yannis and Erik Brynjolfsson. 1999. "Bundling Information Goods: Pricing, Profits, and Efficiency." *Management Science* 45 (12):1613–1630.

Becker, Gary S. 1991. "A Note on Restaurant Pricing and Other Examples of Social Influences on Price." *Journal of Political Economy* 99 (5):1109–1116.

Becker, Gary S. and Kevin M. Murphy. 1993. "A Simple Theory of Advertising as a Good or a Bad." *Quarterly Journal of Economics* 108 (4):941–964.

Borell, C. 1975. "Convex Set Fucntions in d-Space." *Periodica Mathematica Hungarica* 6 (2):111–136.

Bowley, Arthur L. 1924. *The Mathematical Groundwork of Economics*. Oxford: Oxford University Press.

Brander, James A. and Barbara J Spencer. 1981. "Tariffs and the Extraction of Foreign Monopoly Rents under Potential Entry." *The Canadian Journal of Economics* 14 (3):371–389.

Bresnahan, Timothy F. 1989. "Empirical Studies of Industries with Market Power." In *Handbook of Industrial Organization*, vol. 2, edited by Richard Schmalensee and Robert Willig. North-Holland B.V.: Amsterdam, Holland, 1011–1057.

Bulow, Jeremy, John Geanakoplos, and Paul Klemperer. 1985a. "Holding Idle Capacity to Deter Entry." *Economics Journal* 95 (377):178–182.

Bulow, Jeremy I., John D. Geanakoplos, and Paul D. Klemperer. 1985b. "Multimarket Oligopoly: Strategic Substitutes and Compliments." *Journal of Political Economy* 93 (3):488–511.

Bulow, Jeremy I. and Paul Pfleiderer. 1983. "A Note on the Effect of Cost Changes on Prices." *Journal of Political Economy* 91 (1):182–185.

Chen, Yongmin and Michael Riordan. 2008. "Price-Increasing Competition." *RAND Journal of Economics* 39 (4):1042–1058.

Chen, Yongmin and Michael H. Riordan. 2007. "Price and Variety in the Spokes Model." *Economic Journal* 117 (522):897–921.

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods." *Annual Review of Economics* 1:451–488.

Cournot, Antoine A. 1838. *Recherches sur les Principes Mathematiques de la Theorie des Richess*. Paris.

Cowan, Simon. 2010. "Third-Degree Price Discrimination and Consumer Surplus." Http://www.economics.ox.ac.uk/members/simon.cowan/PapersandFiles/ConsumerSurplusResult2.pdf.

Dasgupta, Partha and Eric Maskin. 2000. "Efficient Auctions." *Quarterly Journal of Economics* 115 (2):341–388.

Delipalla, Sofia and Michael Keen. 1992. "The Comparison Between Ad Valorem and Specific Taxation under Imperfect Competition." *Journal of Public Economics* 49 (3):351–367.

Dixit, Avinash and Victor Norman. 1978. "Advertising and Welfare." *Bell Journal of Economics* 9 (1):1–17.

Dixit, Avinash K. and Joseph E. Stiglitz. 1977. "Monopolistic Compeition and Optimum Product Diversity." *American Economic Review* 67 (3):297–308.

Dupuit, Arsène Jules Étienne Juvénal. 1844. *De la Mesure de L'utilité des Travaux Publics*. Paris.

Edelman, Benjamin, Michael Ostrovsky, and Michael Schwarz. 2007. "Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords." *American Economic Review* 97 (1):242–259.

Fabinger, Michal and E. Glen Weyl. 2012. "Pass-Through and Demand Forms." This work is in progress. For a draft of notes, contact Glen Weyl at weyl@uchicago.edu.

Farrell, Joseph. 2008. "Some Welfare Analytics of Aftermarkets." Http://www.law.northwestern.edu/searlecenter/papers/Farrell_Analytics.pdf.

Farrell, Joseph and Carl Shapiro. 1900. "Horizontal Mergers: An Equilibrium Analysis." *American Economic Review* 80 (1):107–126.

———. 2010a. "Antiturst Evaluation of Hoirzontal Mergers: An Economic Alternative to Market Definition." *Berkeley Electronic Journal of Theoretical Ecnomics: Policies and Perspectives* 10 (1).

———. 2010b. "Recapture, Pass-Through and Market Definition." *Antitrust Law Journal* 76 (3):585–604.

Fudenberg, Drew and Jean Tirole. 1984. "The Fat-Cat Effect, the Puppy-Dog Plow, and the Lean and Hungry Look." *American Economic Review* 74 (2):361–366.

Gabaix, Xavier, David Laibson, Deyuan Li, Hongyi Li, Sidney Resnick, and Caspar G. de Vries. 2010. "Extreme Value Theory and the Equilibrium Prices in Large Economies." Http://pages.stern.nyu.edu/∼xgabaix/papers/CompetitionEVT.pdf.

Genesove, David and Wallace P. Mullin. 1998. "Testing Static Oligopoly Models: Conduct and Cost in the Sugar Industry, 1890-1914." *RAND Journal of Economics* 29 (2):355–377.

Gomes, Renato. 2012. "Optimal Auction Design in Two-Sided Markets." Http://sites.google.com/site/northwesternrenatogomes/Home/research.

Goos, Maarten, Patrick Van Cayseele, and Bert Willekens. 2012. "Platform Pricing in Matching Markets." Http://www.econ.kuleuven.be/public/n06022/M2M-20120711-WP.pdf.

Holmes, Thomas J. 1989. "The Effects of Third-Degree Price Discrimination in Oligopoly." *American Economic Review* 79 (1):244–250.

Holmström, Bengt and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7 (Special Issue):24–52.

Ivaldi, Marc, Bruno Jullien, Patrick Rey, Paul Seabright, and Jean Tirole. 2003. "The Economics of Tacit Collusion." Http://idei.fr/doc/wp/2003/tacit_collusion.pdf.

Jaffe, Sonia and E. Glen Weyl. 2010. "Linear Demand Systems are Inconsistent with Discrete Choice." *B. E. Journal of Theoretical Economics (Advances)* 10 (1):Article 52.

———. 2012. "The First-Order Approach to Merger Analysis." Http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1765024.

Jehiel, Philippe and Benny Moldovanu. 2001. "Efficient Design with Interdependent Valuations." *Econometrica* 69 (5):1237–1259.

Jenkin, Henry Charles Fleeming. 1871–1872. "On the Principles which Regulate the Incidence of Taxes." *Proceedings of the Royal Society of Edinburgh* :618–631.

Johnson, Harry G. 1953–1954. "Optimum Tariffs and Retaliation." *Review of Economic Studies* 21 (2):142–153.

Johnson, Justin P. and David P. Myatt. 2006. "On the Simple Economics of Advertising, Marketing, and Product Design." *American Economic Review* 96 (3):756–784.

Klemperer, Paul D. and Margaret A. Meyer. 1989. "Supply Function Equilibria in Oligopoly under Uncertainty." *Econometrica* 57 (6):1243–1277.

Laffont, Jean-Jacques and Jean Tirole. 1993. *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: MIT Press.

Lerner, A. P. 1934. "The Concept of Monopoly and the Measurement of Monopoly Power." *Review of Economic Studies* 1 (3):157–175.

Mankiw, N. Gregory and Michael D. Whinston. 1986. "Free Entry and Social Inefficiency." *RAND Journal of Economics* 17 (1):48–58.

Mares, Vlad and Jeroen M. Swinkels. 2011. "On the Analysis of Asymmertic First Price Auctions." Http://www.kellogg.northwestern.edu/faculty/swinkels/Homepage/Papers/Main%20Paper%205-4-11.pdf.

Marshall, Alfred. 1890. *Principles of Economics*. New York: Macmillan.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6):1695–1725.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon. Forthcoming. “A Reduced-Form Approach to Behavioral Public Finance.” *Annual Review of Economics* .

Nocke, Volker and Michael D. Whinston. Forthcoming. “Merger Policy with Merger Choice.” *American Economic Review* .

Ossa, Ralph. 2011. “A “New Trade” Theory of GATT/WTO Negotiations.” *Journal of Political Economy* 119 (1):122–152.

Perloff, Jeffrey M. and Steven C. Salop. 1985. “Equilibrium with Product Differentiation.” *Review of Economic Studies* 52 (1):107–120.

Pigou, Arthur C. 1920. *The Economics of Welfare*. London: MacMillan.

Prékopa, András. 1971. “Logarithmic Concave Measures with Application to Stochastic Programming.” *Acta Scientiarum Mathematicarum (Szeged)* 32:301–316.

Reny, Philip J., Simon J. Wilkie, and Michael A. Williams. 2012. “Tax Incidence Under Imperfect Competition: Comment.” *International Journal of Industrial Organization* 30 (5):399–402.

Riley, John G. and William F. Samuelson. 1981. “Optimal Auctions.” *American Economic Review* 71 (3):381–392.

Robinson, Joan. 1933. *The Economics of Imperfect Competition*. London: MacMillan.

Rochet, Jean-Charles and Lars A. Stole. 2003. “The Economics of Multidimensional Screening.” In *Advances in Economics and Econometrics: Theory and Applications*, *Eighth World Congress*, vol. 1, edited by Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky. Cambridge University Press, 150–197.

Rochet, Jean-Charles and Jean Tirole. 2003. “Platform Competition in Two-Sided Markets.” *Journal of the European Economic Association* 1 (4):990–1029.

Saez, Emmanuel. 2001. “Using Elasticities to Derive Optimal Income Tax Rates.” *The Review of Economic Studies* 68 (1):205–229.

Schmalensee, Richard. 1981. “Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination.” *American Economic Review* 71 (1):242–247.

Shaked, Avner and John Sutton. 1982. “Relaxing Price Competition Through Product Differentiation.” *Review of Economic Studies* 49 (1):3–13.

Shapiro, Carl. 1996. “Mergers with Differentiated Products.” *Antitrust* 10:23–30.

Spence, A. Michael. 1975. "Monopoly, Quality, and Regulation." *Bell Journal of Economics* 6 (2):417–429.

Spengler, Joseph J. 1950. "Vertical Integration and Antitrust Policy." *Journal of Political Economy* 50 (4):347–352.

Stigler, George J. 1964. "A Theory of Oligopoly." *Journal of Political Economy* 72 (1):44–61.

Telser, Lester G. 1972. *Competition, Collusion and Game Theory*. Chicago: Aldine-Atherton.

Turner, Lesley J. 2012. "The Incidence of Student Financial Aid: Evidence from the Pell Grant Program." Http://www.columbia.edu/∼ljt2110/LTurner_JMP.pdf.

Varian, Hal R. 1985. "Price Discrimination and Social Welfare." *American Economic Review* 75 (4):870–875.

Veiga, André and E. Glen Weyl. 2012. "Multidimensional Product Design." Http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1935912.

Villas-Boas, Sofia Berto and Rebecca Hellerstein. 2006. "Identification of Supply Models of Retailer and Manufacturer Oligopoly Pricing." *Economics Letters* 90 (1):132–140.

von Stackelberg, Heinrich F. 1934. *Marktform und Gleichgwicht*. Vienna: Julius Springer.

Werden, Gregory J. 1996. "A Robust Test for Consumer Welfare Enhancing Mergers Among Sellers of Differentiated Products." *Journal of Industrial Economics* 44 (4):409–413.

Weyl, E. Glen. 2009a. "Monopoly, Ramsey and Lindahl in the Rochet and Tirole (2003)." *Economics Letters* 103 (2):99–100.

———. 2009b. "The Price Theory of Two-Sided Markets." Http://www.people.fas.harvard.edu/∼weyl/research.htm.

Weyl, E. Glen and Jean Tirole. 2012. "Market Power Screens Willingness-to-Pay." *Quarterly Journal of Economics* 127 (4).

Appendix

Our results on deadweight loss assume constant marginal cost, but many of the general intuitions derived there extend to, or are even strengthened with non-linear costs. First, consider the relationship between pass-through and $\frac{dq}{d\tilde{q}}$. Skipping directly to the symmetric imperfect competition model, and not considering the notation-intensive generalization to general imperfect competition,

$$\theta = -\frac{p - mc\left(q - \tilde{q}\right) - t}{p}\frac{p}{\left(q - \tilde{q}\right)p'} = -\frac{p - mc\left(q - \tilde{q}\right) - t}{\left(q - \tilde{q}\right)p'}.$$

However, now the impact of an increase in $\tilde{q}$ directly (the partial derivative) on the right-hand side is

$$-\frac{(q-\tilde{q})\, mc'\,(q-\tilde{q}) + p - mc\,(q-\tilde{q}) - t}{(q-\tilde{q})^2\, p'} = \frac{\theta + \frac{\epsilon_D - \theta}{\epsilon_S}}{q - \tilde{q}}.$$

On the other hand the impact of a change in t is $\frac{1}{(q-\tilde{q})p'}$. Therefore $\frac{dq}{d\tilde{q}} = \left(\theta + \frac{\epsilon_D-\theta}{\epsilon_S}\right)\rho$. Thus when there are declining returns to scale, $\theta\rho$ is smaller and when there are increasing returns $\theta\rho$ is larger than the effect of competition on quantities. Declining returns to scale reduce pass-through and increasing returns increase it, so we can say that returns to scale have a larger impact on $\frac{dq}{d\tilde{q}}$, driving a wedge between them even in the monopoly case of $\theta = 1$.

This competition pass-through, which we now call ρ_c is always below unity if purely demand driven quantity pass-through (that which would prevail with constant returns), $\hat{\rho} \equiv \frac{\theta}{1+\frac{\theta}{\epsilon_\theta}+\frac{\theta}{\epsilon_{ms}}}$ is less than 1. To see this note that

$$\rho_c - 1 = \frac{\theta + \frac{\epsilon_D-\theta}{\epsilon_S}}{1 + \frac{\epsilon_D-\theta}{\epsilon_S} + \frac{\theta}{\epsilon_{ms}} + \frac{\theta}{\epsilon_\theta}} - 1 = \frac{\theta\left(1 - \frac{1}{\epsilon_{ms}} - \frac{1}{\epsilon_\theta}\right) - 1}{1 + \frac{\epsilon_D-\theta}{\epsilon_S} + \frac{\theta}{\epsilon_{ms}} + \frac{\theta}{\epsilon_\theta}},$$

while

$$\hat{\rho} - 1 = \frac{\theta\left(1 - \frac{1}{\epsilon_{ms}} - \frac{\theta}{\epsilon_\theta}\right) - 1}{1 + \frac{\theta}{\epsilon_{ms}} + \frac{\theta}{\epsilon_\theta}} = \frac{\theta\left(1 - \frac{1}{\epsilon_{ms}} - \frac{1}{\epsilon_\theta}\right) - 1}{1 + \frac{\theta}{\epsilon_{ms}} + \frac{\theta}{\epsilon_\theta}}.$$

The numerator of these two expressions is the same and both have positive denominators so long as the equilibrium is stable under constant marginal cost. Thus, assuming such stability, the sign of $\rho_c - 1$ (the strategic effect, substitutes v. complements discussed in Subsection 6.4) is determined by that of $\hat{\rho} - 1$. Notice that decreasing returns to scale move ρ_c towards unity (compared to the constant returns case) by increasing both the numerator and denominator while increasing returns have the opposite effect.

Finally, consider the incidence calculations. For these, it is crucial to specify the costs at which the exogenous units are produced. For this purpose, consider an alternative experiment. Rather than introducing exogenous competition, imagine the state confiscates any profits earned on the first $\underline{q}$ units. Then equilibrium is now

$$\theta = -\frac{p - mc\,(q) - t}{\left(q - \underline{q}\right) p'}$$

and thus $\frac{dq}{d\underline{q}} = \theta\rho$.

Profits per firm are now $p(q)\left(q - \underline{q}\right) - c\,(q) + c\left(\underline{q}\right)$, so the fall in profits from an increase in $\underline{q}$ is

$$p'\theta\rho\left(q - \underline{q}\right) + p\,(\theta\rho - 1) - mc\,(q)\,\theta\rho + mc\left(\underline{q}\right) = m\,(-\rho + \theta\rho - 1 + \alpha) = -m\left[1 + \rho\,(1-\theta) - \alpha\right].$$

where $\alpha \equiv \frac{mc(q)-mc(\underline{q})}{m}$. The argument for calculating deadweight loss incidence proceeds exactly as before, so we obtain relative efficiency gains compared to profit losses of

$$\frac{\theta\rho}{1+\rho(1-\theta)-\alpha}.$$

These can be converted, just as in the text, to global incidence formulas. Notice that with decreasing returns, $\alpha > 0$ and thus deadweight loss is larger relative to profit than given by the formula in the text. This is intuitive because with decreasing returns the existence of a competitive rent makes profits positive even in the absence of a monopoly distortion. When returns are increasing $\alpha < 0$ and thus deadweight loss is larger relative to profits than the formula given in the text indicates. Again this is intuitive as with increasing returns the competitive rent is negative, reducing profits relative to deadweight loss. Thus the basic source of the divergence is that while in the main text changes in cost only affected profits, here increasing costs also affect the size of the deadweight loss triangle directly. Thus increasing or decreasing marginal costs have an additional impact on the $\frac{DWL}{PS}$ that they do not have on the $\frac{CS}{PS}$ ratio.